H. FERNÁNDEZ-ABASCAL,
MARTA M. GUIJARRO,
JOSÉ LUIS ROJO Y JOSÉ A. SANZ

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA

Ariel Economía

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA

ARIEL ECONOMÍA

H. FERNÁNDEZ-ABASCAL, MARTA M. GUIJARRO,
JOSÉ LUIS ROJO y JOSÉ A. SANZ

CÁLCULO DE PROBABILIDADES Y ESTADÍSTICA

EDITORIAL ARIEL, S. A.
BARCELONA

1.ª edición: junio 1994

© 1994: H. Fernández-Abascal, Marta M. Guijarro, José Luis Rojo y José A. Sanz

Derechos exclusivos de edición en castellano
reservados para todo el mundo:
© 1994: Editorial Ariel, S. A.
Córcega, 270 - 08008 Barcelona

ISBN: 84-344-2094-5

Depósito legal: B. 20.664 - 1994

Impreso en España

Ninguna parte de esta publicación, incluido el diseño de la cubierta, puede ser reproducida, almacenada o transmitida en manera alguna ni por ningún medio, ya sea eléctrico, químico, mecánico, óptico, de grabación o de fotocopia, sin permiso previo del editor.

PREFACIO

El libro que se presenta a continuación recoge los conceptos básicos del Cálculo de Probabilidades y de las distribuciones de probabilidad. Es el primero de una serie de dos, que se completará con otro dedicado al estudio de la Inferencia Estadística.

Su contenido incluye los desarrollos y aplicaciones más usuales del Cálculo de Probabilidades que son de interés para las ciencias aplicadas. Se aborda, de este modo, la construcción axiomática de la teoría de la probabilidad en los capítulos segundo y tercero, dedicando una especial atención al manejo de las herramientas de cálculo habituales (probabilidad condicionada, independencia de sucesos, teorema de la probabilidad total, teorema de Bayes, composición de experimentos, etc.).

En los capítulos comprendidos entre el cuarto y el octavo se estudian las variables aleatorias. Los dos primeros se dedican a desarrollar los conceptos básicos del Cálculo de Probabilidades con variables (probabilidad inducida, función de distribución, etc.), prestando una gran atención a las variables aleatorias discretas y a las continuas, que aparecen en la casi totalidad de las aplicaciones en las que intervienen variables aleatorias. Se realiza un desarrollo especialmente cuidadoso de las distribuciones condicionadas y del concepto de independencia de variables aleatorias, de especial interés en la formalización de modelos probabilísticos complejos en el campo aplicado.

En los capítulos sexto y séptimo se desarrollan las características numéricas asociadas a las variables aleatorias, a saber, esperanzas o medias, medianas, varianzas, coeficientes de asimetría y de aplastamiento, covarianzas, coeficientes de correlación, etc. Una buena parte del capítulo séptimo se dedica a la regresión, tanto en su enfoque lineal como en el general. La función característica, una característica de las variables aleatorias de difícil comprensión, aunque de fácil manejo para estudiar el comportamiento probabilístico de sumas y medias, ocupa el capítulo octavo de este texto.

Los capítulos que van del noveno al duodécimo muestran un repertorio de las distribuciones de probabilidad más usuales en la modelización de fenómenos aleatorios. Para cada una de las distribuciones descritas se calculan sus características numéricas más importantes y se enuncian y demuestran las propiedades más usuales; esta parte se completa con un anexo situado al final del libro, que incluye las tablas estadísticas de las principales distribuciones, tablas que permiten el cálculo aproximado y rápido de probabilidades.

El capítulo decimotercero, y último, estudia las convergencias de sucesiones de variables aleatorias, herramienta imprescindible para obtener distribuciones de probabilidad aproximadas para las sumas y medias de muchas variables aleatorias a través del teorema del límite central y de las denominadas leyes de los grandes números. Sobre estos últimos resultados descansa una buena parte de la Inferencia Estadística, a la que dedicaremos el segundo de los libros.

Para el estudio completo del libro se precisan unos conocimientos moderados de Cálculo Diferencial e Integral, algunas nociones del cálculo de límites y de las técnicas de sumación de series numéricas sencillas, así como ciertos conceptos y desarrollos de álgebra lineal. Es necesario manejar con soltura el Análisis Combinatorio, aunque el apéndice que figura al final del capítulo segundo recuerda las motivaciones estadísticas y los resultados fundamentales relacionados con el mismo.

Cada capítulo admite varias lecturas, dependiendo del nivel de preparación del lector. En general, las definiciones se proporcionan con el necesario rigor, y se realizan las demostraciones de los resultados siempre que su seguimiento sea posible para un lector que se encuentre en las condiciones especificadas en el párrafo anterior. No obstante, las definiciones y resultados básicos (y muchos no tan básicos) van acompañados de observaciones que permiten que el lector poco avezado en las lides matemáticas pueda valorar el contenido de los mismos, haciéndose, al menos, una idea intuitiva de ellos. En este mismo sentido, tanto las definiciones como los resultados básicos mencionados anteriormente van acompañados de multitud de ejemplos que los ilustran. Cada capítulo finaliza con una amplia colección de ejercicios propuestos, a través de los cuales el lector podrá medir el grado de comprensión de los contenidos de cada capítulo.

El libro está escrito para que sirva de manual y de obra de consulta para un curso cuatrimestral de Cálculo de Probabilidades, estando orientado esencialmente a los alumnos de los nuevos planes de estudio de las licenciaturas y diplomaturas en Economía y en Administración de Empresas de las Facultades y Escuelas Universitarias de Ciencias Económicas y Empresariales. Pero puede utilizarse con el mismo provecho, si bien con una lectura diferente, en licenciaturas y diplomaturas que necesiten de un bagaje adecuado de técnicas relacionadas con el Cálculo de Probabilidades (estadística, ciencias físicas, químicas y biológicas, ciencias sociales y humanas o ciencias médicas, entre otras). Esta multiplicidad de posibilidades, así como el abultado número de páginas del libro, es consecuencia del deseo de los autores de suministrar al alumno de esas licenciaturas y diplomaturas un texto que no se le quede corto cuando desarrolle su andadura profesional, al menos en cuanto a los desarrollos más generales.

Los autores poseen una larga trayectoria docente en el campo de la economía aplicada, en concreto, en estadística y econometría, en las Facultades de Ciencias Económicas y Empresariales. Hermenegildo Fernández-Abascal Teira es profesor titular de Escuela Universitaria, José Luis Rojo García es catedrático de Universidad, y José Antonio Sanz Gómez es profesor asociado, todos ellos en el Departamento de Economía Aplicada (Estadística y Econometría) de la Universidad de Valladolid. Marta Guijarro Garvi ha sido profesora asociada en este mismo Departamento, y en la actualidad es profesora titular de Universidad en el Departamento de Economía de la Universidad de Cantabria.

El libro que llega a las manos del lector es el resultado de una larga reflexión colectiva, desarrollada durante los últimos dieciséis años, de la que han participado, en diferentes períodos, diversos profesores de Estadística y Econometría de la Facultad de Ciencias Económicas y Empresariales de Valladolid. A todos ellos, así como a los alumnos de las distintas promociones que nos han premiado con su entendimiento y paciencia, y a veces con su incomprensión, nuestro agradecimiento.

Valladolid y Santander, febrero de 1994.

Capítulo 1

INTRODUCCIÓN

Antes de abordar el estudio de cualquier disciplina, parece obligado preguntarse sobre su naturaleza, su objeto y sus métodos. Aunque muchas veces resulte retórico, cuando no estéril, tratar de delimitar todo un mundo entre unas escasas palabras, una definición, siempre provisional y sujeta a revisión, puede ser un buen punto de partida para obtener una primera visión totalizadora de la materia en cuestión.

A la pregunta «**¿qué es la Estadística?**» podemos dar múltiples respuestas, pero es seguro que en cada una de ellas dejaríamos escapar parte de su misma esencia, mientras que incluiríamos ciertos contenidos que no le son propios. De entrada podemos encontrar dos acepciones diferentes de la palabra «estadística» .

En los medios de comunicación y en la calle oímos hablar de las estadísticas del paro, de las estadísticas de accidentes de tráfico o de las estadísticas de divorcios. En todos estos casos, el término **estadísticas** (así, en minúscula y normalmente en plural) se refiere a una colección numérica de datos.

Pero en esos mismos medios también oímos hablar de que la Estadística confirma la mayor incidencia del cáncer de pulmón entre los fumadores o de técnicas estadísticas que han ayudado a diseñar la campaña de publicidad para el lanzamiento de un nuevo automóvil. En esta acepción se engloban tanto el denominado Cálculo de Probabilidades como la **Estadística** (ahora sí, con mayúscula), una rama científica relativamente joven, pero con un fructífero desarrollo a lo largo de este siglo. De esta segunda acepción trataremos a lo largo del presente capítulo, refiriéndonos con el nombre de Estadística tanto al Cálculo de Probabilidades como a la Estadística propiamente dicha.

En primer lugar, trataremos de acercarnos al verdadero concepto de la Estadística con la vana intención de poder llegar a dar una definición (sección 1.1). Más adelante, situaremos la Estadística entre las ramas de las matemáticas, reflexionando sobre los métodos propios de esta ciencia formal (secciones 1.2 y 1.3). Una breve referencia a la historia de la Estadística nunca será un recurso erudito; el conocimiento de su pasado nos servirá para aproximarnos más a su naturaleza y nos proporcionará las claves de lo que puede ser su futuro (sección 1.4). Por último, haremos unas consideraciones sobre su carácter de ciencia instrumental. La Estadística ha experimentado un gran desarrollo debido fundamentalmente a su

función de apoyo a otras ciencias, tanto a aquellas que tienen por objeto la naturaleza inanimada (Física, Química, Astronomía, etc.) y la naturaleza animada (Medicina, Biología, Zoología, etc.) como a aquellas que consideran el entorno del hombre desde su perspectiva social (Economía, Sociología, etc., es decir, las llamadas Ciencias Sociales). Haremos especial mención a la importancia de la utilización de la Estadística en el ámbito de estas ciencias (sección 1.5).

1.1 Una aproximación al concepto de Estadística

De cara a tener un punto de referencia vamos a intentar, en primer lugar, encontrar una definición, más o menos satisfactoria, de la Ciencia Estadística.

Antes de nada, actuemos de la forma más sensata en estos casos; es decir, consultemos un diccionario. Con las siguientes palabras María Moliner[1] trata de definir el vocablo estadística:

> *Ciencia que se ocupa del recuento de hechos sociales, científicos o de cualquier clase y de la comparación de las cifras obtenidas.*

Como muchas de las definiciones «de diccionario» , su brevedad y generalidad hacen que se escape parte de la esencia de lo definido. La definición anterior, válida para una parte de la Estadística, la Descriptiva, no recoge la naturaleza y sentido de la otra parte, la Estadística Teórica o Matemática.

Sigamos la búsqueda de esa definición por el mismo camino y acudamos a un diccionario enciclopédico que, por su propio carácter, no está sujeto a la brevedad y concisión requeridas en un diccionario de la lengua. Así, la *Gran Enciclopedia Larousse*[2] recoge tres acepciones de la palabra estadística:

1. *Conjunto de datos de observación relativos a un grupo de individuos o de unidades (suele usarse en plural).*
2. *Conjunto de los métodos que tienen por objeto la obtención, el tratamiento y la interpretación de dichos datos (estas dos últimas actividades, así como el paso a leyes y a modelos teóricos capaces de representar esos datos, competen a la Estadística matemática).*
3. *Conjunto de datos numéricos que se refieren a un fenómeno cualquiera y del que se extraen determinadas conclusiones.*

La segunda acepción propuesta resulta una buena definición en cuanto que recoge las principales fases de la metodología estadística, haciendo mención a los dos enfoques: el descriptivo (sin nombrarlo) y el matemático o teórico (éste explícitamente). No es la única virtualidad de esta definición; al referirse a la primera acepción pone en evidencia la materia prima de la Estadística: los datos u observaciones.

1. MOLINER, M.: *Diccionario de uso del español*, 2 vols., Madrid, Gredos, 1991, p. 1223.
2. *Gran Enciclopedia Larousse*, 24 vols., Barcelona, Planeta, 1990, vol. 9, p. 3985.

Otra posibilidad que se nos presenta para captar el concepto de Estadística es buscar su origen etimológico. La palabra deriva del término latino *status*, estado, y aparece por primera vez en los escritos de G. ACHENWALL. A lo largo de los siglos XVII y XVIII se desarrolla principalmente en Alemania un enfoque de la Estadística que se materializa en la llamada Escuela Descriptiva, que tiene por objeto el estudio de lo que el Estado tiene de cuantificable (lo que hoy entendemos por Geografía Política). En un libro de uno de sus componentes, G. ACHENWALL, publicado en Londres en 1770, aparece por primera vez la palabra estadística entendida como *la exteriorización cuantitativa de las cosas del Estado*[3].

Para completar este mosaico de definiciones, oigamos a los hombres que *hacen* estadística. Por un lado M. G. KENDALL[4] afirma:

> *La estadística es la matriz de toda ciencia experimental, y, por consiguiente, una rama del método científico, si no el método científico por excelencia.*

W. H. KRUSKAL[5] va más allá y afirma que *la estadística es parte de la filosofía de la ciencia.*

Otras dos definiciones introducirán un nuevo e importante concepto: la incertidumbre. Así, A. MOOD y F. A. GRAYBILL[6] afirman que

> *La estadística proporciona instrumentos para la toma de decisiones cuando prevalecen condiciones de incertidumbre.*

Remachando lo anterior M. H. DEGROOT[7] dice al respecto:

> *La ciencia estadística se ocupa del desarrollo de teorías y técnicas apropiadas para realizar inferencias bajo condiciones de incertidumbre e ignorancia parcial que necesariamente existen en una amplia gama de actividades. La práctica estadística corriente depende fundamentalmente de la formulación de modelos probabilísticos de diversos sistemas físicos, de métodos de recogida y análisis de datos numéricos y del diseño de experimentos eficientes e informativos.*

Todas y cada una de las definiciones anteriores han aportado una cierta visión del concepto de Estadística, pero la misma complejidad que muestran indica la dificultad de encontrar una única definición totalizadora.

3. MARTÍN-GUZMÁN, M. P. Y MARTÍN PLIEGO, F. J.: *Curso básico de estadística económica*, Madrid, AC, 1989, p. 1.
4. KENDALL, M. G.: «Sobre el futuro de la Estadística. Una segunda visión» , Madrid, *Estadística Española*, núm. 41, 1968, p. 7.
5. KRUSKAL, W. H.: *Enciclopedia internacional de las Ciencias Sociales*, Madrid, Aguilar, 1979, vol. 4, p. 398.
6. MOOD, A. Y GRAYBILL, F. A.: *Introducción a la Teoría de la Estadística*, Madrid, Aguilar, 1978, p. 3.
7. DEGROOT, M. H.: *Optimal Statistical Decisions*, New York, McGraw-Hill, 1970, p. 3.

Desde la Antigüedad se han desarrollado dos concepciones del mundo contrapuestas. Por un lado, existe una concepción determinista que entiende la realidad en base a una continua concatenación de causas y efectos; la labor del científico, en cualquiera de las especialidades, consistiría en desentrañar esas relaciones causales. Utilizando el símil que describe la realidad como un inmenso reloj donde infinitas ruedecillas van proporcionándose movimiento unas a otras, el científico sería el relojero capaz de comprender la maquinaria, pudiendo prever sus movimientos futuros e, incluso, pudiendo alterarlos para conseguir un cierto fin. Todas las ciencias, en menor o mayor medida, son deudoras de esa visión determinista y, por tanto, explicable de la realidad. La otra concepción, la aleatoria, entiende el mundo como fruto del azar; ya no hay un reloj para que el científico desentrañe su mecanismo. En última instancia todo iría abocado al desorden y al caos. Las ciencias tradicionales poco tienen que decir ante esta realidad caprichosa e inexplicable. Si aún así, dentro de ese desorden pudiésemos atisbar una cierta lógica del azar, necesitaríamos de una herramienta para movernos en un mundo que no admite verdades absolutas; esa herramienta sería la Ciencia Estadística.

Hoy, a finales del siglo XX, las reflexiones de un científico no suelen ir por estos caminos. Tal vez la realidad no sea aleatoria, sino determinista, pero, en todo caso, eso importa poco al científico. A. EINSTEIN decía que él no creía que Dios jugase a los dados, lo que ocurre es que a veces es conveniente admitir esa afición divina por el juego.

Quizás el mundo pueda entenderse como ese gran reloj al que nos referíamos. Ahora bien, la complejidad de su mecanismo y la limitación de nuestros medios y conocimientos hacen rehuir al científico de la explicación total de sus movimientos. Fijará su interés solamente en las causas y efectos principales de la realidad a estudiar, renunciando, por economía científica, ignorancia o imposibilidad, a la explicación de ciertas causas y efectos secundarios y no fundamentales. Uno de los padres del Cálculo de Probabilidades y, por otra parte, determinista convencido, P. S. DE LAPLACE [8] escribía a propósito:

> *Una inteligencia que en un momento determinado conociera todas las fuerzas que animan a la naturaleza, así como la situación respectiva de los seres que la componen, si además fuera lo suficientemente amplia como para someter a análisis tales datos, podría abarcar en una sola fórmula los movimientos de los cuerpos más grandes del universo y los del átomo más ligero; nada le resultaría incierto y tanto el futuro como el pasado estarían presentes ante sus ojos. El espíritu humano ofrece, en la perfección que ha sabido dar a la astronomía, un débil esbozo de esta inteligencia. Sus descubrimientos en mecánica y geometría, junto con el de la gravitación universal, le han puesto en condiciones de abarcar en las mismas expresiones analíticas los estados pasados y futuros del sistema del mundo. Aplicando el mismo método a algunos otros objetos de su conocimiento, ha logrado reducir*

8. LAPLACE, P. S. DE: *Ensayo filosófico sobre las probabilidades*, Madrid, Alianza, 1985, pp. 25-26.

> *a leyes generales los fenómenos observados y a prever aquellos otros que deben producirse en ciertas circunstancias. Todos sus esfuerzos por buscar la verdad tienden a aproximarlo continuamente a la inteligencia que acabamos de imaginar, pero de la que siempre permanecerá infinitamente alejado. Esta tendencia, propia de la especie humana, es la que la hace superior a los animales, y sus progresos en este ámbito, lo que distingue a las naciones y los siglos y cimenta su verdadera gloria.*

En este mismo sentido LAPLACE[9] afirma:

> *Para el hombre hay por tanto muchas cosas que son inciertas y algunas que son más o menos probables. En vista de la imposibilidad de conocerlas todas, ha tratado de compensar esto determinando distintos grados de apariencia, de suerte que debemos a la debilidad de la mente humana una de las más delicadas e ingeniosas teorías matemáticas: la ciencia del azar o probabilidad.*

Bien sea por la naturaleza de la realidad, bien por las características de nuestro conocimiento, la utilización de la Estadística en todos los campos científicos es una herramienta imprescindible a la que no se puede renunciar, por lo que la discusión anterior sobre la aleatoriedad no debe conducir al científico a posturas nihilistas.

Todo ello nos sugiere una nueva definición de Estadística. Entenderemos un **experimento determinista** como aquel que realizado bajo las mismas condiciones produce los mismos resultados; es decir, una misma causa provoca un mismo efecto. Los experimentos deterministas ponen en evidencia fenómenos deterministas. Las ciencias experimentales (física y química, principalmente) nos proporcionan múltiples ejemplos de estas situaciones. Por otro lado, consideraremos un **experimento aleatorio** aquel que realizado bajo las mismas condiciones produce distintos e impredecibles resultados; es decir, se rompe la relación unívoca causa-efecto. Los fenómenos aleatorios son aquellos que subyacen a este tipo de experimentos.

En esta definición de experimento y fenómeno determinista y aleatorio obviamos las consideraciones anteriores sobre el carácter determinista o indeterminista de la realidad o de nuestro conocimiento. El ejemplo clásico del lanzamiento de un dado nos lo revela. ¿Es realmente aleatorio el resultado obtenido? O más bien, ¿es fruto de nuestra ignorancia y pereza a la hora de desentrañar todas las circunstancias físicas determinantes de dicho lanzamiento? Sean cuales sean las respuestas a dichas preguntas, nuestra consideración sobre la aleatoriedad del fenómeno no variará. Téngase en cuenta la simplificación que supone hablar de repetición del experimento bajo las mismas condiciones. La mayoría de las situaciones, y fundamentalmente las que a nosotros nos interesan, no admiten esta consideración. Aun así, nosotros mantenemos esas premisas de igualdad de condiciones, siendo las desviaciones respecto a esas condiciones ideales una fuente más de aleatoriedad.

9. LAPLACE, P. S. DE: *Oeuvres Complètes*, 14 vols., París, Courcier, 1878 y 1912, vol. 8, p. 114.

De todo ello podemos extraer una nueva y tal vez más precisa definición: **la Estadística es la ciencia que se ocupa del estudio de los fenómenos aleatorios.**

La Estadística presenta dos planteamientos claramente diferenciados; por un lado, uno simplemente descriptivo y empírico y, por otro, un planteamiento más trascendente y teórico.

La **Estadística Descriptiva** no saca conclusiones que vayan más allá de los propios datos que maneja. A partir de las observaciones, más o menos numerosas, las ordena, las resume y las analiza, extrayendo sólo la información primaria contenida en las mismas. Dicha rama de la Estadística está constituida por un conjunto de técnicas y métodos para la recogida, descripción y análisis de datos, tendentes a obtener una serie de resultados, que en cierta medida resuman la información. Su labor ante los fenómenos aleatorios es estudiarlos una vez han ocurrido; por tanto, se trata de un estudio empírico, *a posteriori*, basado en la ocurrencia observada de los datos (frecuencias).

La **Estadística Teórica** (también llamada Matemática o Inferencial) tiene por objeto el estudio de los fenómenos aleatorios antes de que éstos se produzcan; de cara a encontrar en ellos unas pautas de comportamiento, es decir, de cara a preverlos. Es por tanto un estudio teórico, *a priori*, basado en la ocurrencia esperada de los posibles resultados (probabilidades).

Estos dos planteamientos no son antagónicos. Para prever las realidades aleatorias tenemos que basarnos en su historia, en observaciones pasadas de la misma. Es decir, para hacer Estadística Teórica tenemos que apoyarnos en la Estadística Descriptiva que nos proporcionará resúmenes del comportamiento pasado de los fenómenos aleatorios que tratamos de estudiar.

La probabilidad, como medida teórica de la ocurrencia de los resultados de un experimento aleatorio, admite diversas concepciones que se han desarrollado a lo largo de su historia y que expondremos esquemáticamente.

1) *Interpretación laplaciana.* Basándose fundamentalmente en las situaciones generadas por los juegos de azar, LAPLACE[10] define la probabilidad de un suceso como *la razón entre el número de casos favorables y el de todos los casos posibles.* Como veremos en la sección 2.5.1, esta definición tiene validez únicamente cuando consideramos un número finito de resultados posibles y todos ellos igualmente probables.

2) *Interpretación frecuencialista.* Se fundamenta en la ley empírica de regularidad estadística: la frecuencia relativa de un suceso tiende a estabilizarse cuando el número de experimentos crece indefinidamente. Entonces definiremos la probabilidad como el número al que *tiende* la frecuencia relativa, es decir, se define como el límite (no matemático) de la frecuencia

$$pr(A) = \lim fr(A) .$$

10. LAPLACE, P. S. DE: *Ensayo...*, *op. cit.*, p. 31.

Esta interpretación, que aparece en la literatura estadística a finales del siglo pasado y que analizaremos en la sección 2.3, es formalizada por R. VON MISES en 1920, y servirá de motivación para introducir la construcción axiomática de la probabilidad propuesta por A. KOLMOGOROV trece años después.

3) *Interpretación subjetivista o personalista.* En este caso, la probabilidad de un suceso se interpreta como el grado de creencia que para una persona determinada tiene, en base a una información *a priori* sobre el mismo. Este grado de creencia sería distinto según la persona que lo mantuviese, pero tenderá a unificarse en la medida en que la información sobre el mismo se vaya completando. La forma de cuantificar esta probabilidad se hace en base a la idea de plantearse una apuesta sobre la ocurrencia del suceso en cuestión y viendo el límite que una persona estaría dispuesta a apostar.

Esta interpretación, intuida por E. BOREL y B. DE FINETTI en los años veinte, es sistematizada por L. I. SAVAGE en 1954.

El enfoque bayesiano del Cálculo de Probabilidades recoge esta interpretación subjetivista, incorporando en distintas etapas información sobre el suceso que va *mejorando* la cuantificación de la probabilidad del mismo.

4) *Interpretación lógica.* Introducida en 1921 por el economista J. M. KEYNES en su libro *A treatise on probability*, se considera la probabilidad como una relación entre una proposición y un conjunto de ellas que representan la evidencia dada. En la medida en que varíen las evidencias cambiará la probabilidad de la proposición, pero a diferencia del enfoque subjetivista este grado de creencia no depende de los individuos. H. JEFFREYS, en los años sesenta, amplió y profundizó en esta teoría.

Estos enfoques complementarios no influyen en el planteamiento de un curso básico de Cálculo de Probabilidades y Estadística como el que aquí se pretende.

1.2 El método estadístico

Analicemos a continuación las distintas fases que se siguen al aplicar el método estadístico.

> *La investigación científica es un proceso de aprendizaje dirigido. El objeto de los métodos estadísticos es hacer que ese proceso sea lo más eficiente posible. El aprendizaje* ... [avanza en un esquema iterativo en el que] ... *una hipótesis inicial conduce, por un proceso de deducción, a ciertas consecuencias necesarias que pueden ser comparadas con datos. Cuando las consecuencias y los datos no concuerdan la discrepancia puede conducir, por un proceso denominado inducción, a la modificación de la hipótesis. Se inicia entonces un segundo ciclo de iteración. Se deducen las consecuencias de la hipótesis modificada y se comparan de nuevo con los datos (los que ya teníamos o nuevos; en*

> *la mayoría de nuestras aplicaciones el proceso de obtención de los datos es por medio de la experiencia científica, pero pueden obtenerse de una biblioteca desenterrando información ya existente o bien mediante una encuesta) que a su vez pueden llevar a nuevas modificaciones y ganancia de conocimiento.*[11]

Las etapas de aplicación de los métodos estadísticos, que resumiremos a continuación, no coinciden con el orden de los capítulos de un manual de Cálculo de Probabilidades y Estadística. En concreto, se desarrollan en las siguientes fases:

1) *Planteamiento del problema.* Tenemos que limitar el problema en cuestión a términos abordables, concretando su alcance, su ámbito de aplicación (población y unidades sobre las que se va a operar) y las características a tener en cuenta (variables a observar). Esta fase concluye con la propuesta de un método de análisis del problema.

2) *Obtención de datos.* Mediante distintos métodos recogeremos la información requerida sobre la población fijada. Las dos maneras básicas de realizar las observaciones son:

a) Diseño de experimentos.

b) Encuestación (muestral o censal).

3) *Análisis descriptivo.* Presentaremos los datos en cuadros, tablas y gráficos que contengan toda la información ya ordenada. Para comprender mejor la información se resumirán ciertos aspectos de la misma con alguna medida apropiada. Nos valdremos de la Estadística Descriptiva para alcanzar los objetivos de esta etapa.

4) *Análisis teórico.* Se intenta explicar los datos obtenidos mediante un apropiado modelo estadístico, estimando parámetros y contrastando hipótesis; en fin, validando o invalidando el modelo. Esta etapa se cubre con la Estadística Teórica en algunas de sus múltiples ramas.

5) *Explotación del modelo.* Si los datos no invalidan el modelo propuesto, se procede a su explotación. En esta fase, el Cálculo de Probabilidades juega un papel relevante.

1.3 La Estadística como rama de las matemáticas

El Cálculo de Probabilidades y la Estadística son ramas de las matemáticas al igual que la Aritmética, el Álgebra o la Geometría. Cualquier rama de las matemáticas no es otra cosa que el estudio de un modelo matemático, es decir,

11. BOX, G. Y OTROS: *Estadística para investigadores*, Barcelona, Reverté, 1988, pp. 1-2.

de *una representación abstracta simplificada de un cierto fenómeno real*[12]. En palabras de ALEKSANDROV[13], ambas comparten con el resto de las ramas *ciertos rasgos característicos: su abstracción, su precisión, su rigor lógico, el irrefutable carácter de sus conclusiones y, finalmente, el campo excepcionalmente amplio de sus aplicaciones.*

Veamos de forma sucinta cómo se construye el modelo matemático. Fijémonos para ello en una de las ramas más antiguas y a la vez más presente en sus aplicaciones: la Geometría. Esta rama centra su estudio en el espacio físico en que nos movemos, si bien posteriores generalizaciones han abierto sus miras a espacios no reales pero con un campo de aplicación inmenso (en 1826, N. I. LOBACHEVSKI desarrolló una teoría geométrica en base a la negación del quinto postulado de EUCLIDES; años después, esta Geometría, que sólo tenía en principio un interés académico, sirvió de base para la formalización de las teorías de la relatividad, entre otras). La Geometría tiene su origen en la necesidad de medir las tierras (ése es el significado etimológico de la palabra *geometría*); las continuas inundaciones en Egipto, provocadas por las crecidas del Nilo, borraban los límites de los campos y la restitución de dichos límites exigía unas mediciones exactas y objetivables. En el proceso de construcción de la Geometría se extrae de la realidad aquello que interesa estudiar prescindiendo del resto. Así, ante la contemplación en la naturaleza de objetos con líneas rectas (la cuerda tensa de un arco, la plomada en la construcción de edificios, la línea del horizonte, etc.) el hombre intuye la idea de recta y la conceptualiza quitando todo lo que le es ajeno, la inmaterializa. De esta misma forma fue definiendo el hombre los distintos elementos, punto, recta, plano, etc., y figuras geométricas (lo que constituyen las definiciones) y enunciando relaciones que los ligaban (axiomas). Mediante razonamientos lógicos dedujo teoremas (el de Tales, el de Pitágoras, etc.) y propiedades, cuyos resultados aplicó para medir, por ejemplo, un terreno rectangular (pero no uno sólo, sino cualquier terreno de esas características).

Lo anterior pone en evidencia dos características comunes a cualquier rama matemática:

- El origen y desarrollo de cualquiera de ellas se encuentra en las necesidades y exigencias del mundo que nos rodea.
- Su construcción es axiomática.

El método axiomático del que participan el Cálculo de Probabilidades y la Estadística se desarrolla en varias etapas:

1) Se parte de una realidad, uno de cuyos aspectos (en la Geometría el espacio físico) se quiere estudiar. De esta realidad se extraen una serie de intuiciones sobre los elementos y sus relaciones.

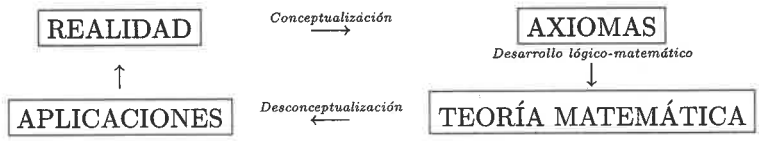
12. RÍOS, S.: *Métodos Estadísticos*, Madrid, Castillo, 1975, p. 19.
13. ALEKSANDROV, A. D. Y OTROS: *La matemática: su contenido, métodos y significado*, Madrid, Alianza, 1976, p. 17.

2) Se enuncian una serie de definiciones (abstracción de los elementos, prescindiendo de lo superfluo) y axiomas (las relaciones ideales entre esos elementos). Los axiomas son por naturaleza no demostrables, compatibles (no contradictorios) y, a poder ser, independientes.

3) A partir de las definiciones y axiomas y mediante razonamientos lógicos, básicamente deductivos, se obtienen una serie de teoremas, proposiciones, corolarios, y propiedades.

4) Dichos resultados, desde su generalidad, los podemos aplicar a la realidad de la que partíamos, satisfaciendo algunas de las necesidades que había generado la correspondiente Teoría Matemática.

El esquema siguiente, tomado de *Métodos Estadísticos* de S. RIOS[14], describe y resume con claridad lo expuesto.



La ventaja que trae consigo esta forma de trabajar es grande. Al estudiar modelos ideales estudiamos situaciones comunes a muy diversos fenómenos, pudiéndose aplicar los resultados obtenidos a distintas realidades concretas, con la consiguiente economía de estudio. La realidad siempre es más rica que el modelo, pero estudiando lo que tienen de común distintas realidades conseguimos un conocimiento general mucho más profundo. La validez del modelo planteado depende de lo realista que éste sea. Por tanto se establece un compromiso entre la abstracción y el realismo de un modelo: el modelo ha de ser lo suficientemente ideal para abarcar muchos casos concretos pero, también, lo suficientemente real para no alejarse en demasía de cada uno de dichos casos.

Así como la Geometría tiene por objeto de estudio el espacio físico, el objeto de la Estadística son los fenómenos aleatorios. Mientras que las primeras construcciones axiomáticas de la Geometría se remontan al siglo III antes de Cristo, la axiomatización de la Estadística tiene poco más de medio siglo.

En 1933, A. KOLMOGOROV en sus *Fundamentos de la Teoría de la Probabilidad*[15], propuso una serie de definiciones (sucesos elementales, suceso, σ-álgebra, etc.) y axiomas (véase la sección 2.3) que pusieron los cimientos del Cálculo de Probabilidades y por ende de la Estadística. Esta axiomatización ya había sido intuida por E. BOREL años antes, ligando el concepto de probabilidad a la ya desarrollada Teoría de la Medida.

De esto no hay que deducir que toda la Estadística y el Cálculo de Probabilidades anterior a 1933 carezca de valor y sea desechable. Al definir axiomáticamente la probabilidad se consigue encajar todas las piezas anteriores como si fuese un

14. *Op. cit.*, p. 19.
15. KOLMOGOROV, A.: *Foundations of the Theory of Probability*, New York, Chelsea, 1956, p. 2.

inmenso rompecabezas y se sientan las bases de un crecimiento más rápido y lógico de nuevos métodos estadísticos.

La Estadística también comparte con el resto de ramas de las matemáticas una forma común de razonamiento: el *método deductivo*. La deducción es la forma de razonamiento que parte de un principio general conocido para llegar a un principio particular desconocido. En algunos casos, deducciones constructivas, la conclusión contiene algún elemento nuevo que no estaba presente en las premisas. Según cuál sea el punto de partida, las premisas o las conclusiones, se habla de deducción progresiva o regresiva. Esta última es habitual en matemáticas, donde ciertos principios particulares han hecho intuir principios más generales, demostrados posteriormente.

1.4 Desarrollo histórico del Cálculo de Probabilidades y de la Estadística

Una breve reseña sobre la historia de cualquier disciplina es siempre enriquecedor desde muy diversas perspectivas; y mucho más en una materia como la Estadística que comparte las abstracciones más elevadas de las ciencias matemáticas (Estadística teórica) con el apego a la más inmediata realidad (Estadística aplicada).

Según D. PEÑA[16]:

> *... el conocimiento de la historia de una disciplina es importante, al menos en tres aspectos: el primero, para entender su estado actual de desarrollo y la relación entre sus partes; el segundo, para comprender su terminología, ya que el nombre de una técnica o de un método suele estar asociado a sus orígenes históricos; el tercero, para prever su desarrollo futuro.*

A esto habría que añadir que el conocimiento de la génesis de los principales conceptos estadísticos puede servir para la buena comprensión de los mismos. Bien es verdad que las características de un libro como éste no aconsejan más que un breve esbozo de esta apasionante historia.

La dificultad de resumir en unas pocas páginas la historia de toda una ciencia, aunque ésta, en sentido estricto, sólo abarque un par de siglos, es patente como se manifiesta en las palabras de M. G. KENDALL[17]:

> *No es posible señalar una fuente única e identificable del ancho río de pensamiento que se conoce hoy por el nombre de estadística teórica. Ésta es más bien la confluencia, a lo largo de más de dos siglos, de cierto número de corrientes tributarias procedentes de muchas regiones*

16. PEÑA, D.: *Estadística, Modelos y métodos*, Madrid, Alianza, 1986, vol. 1, pp. 32 y 34.
17. KENDALL, M. G.: *Enciclopedia Internacional de las Ciencias Sociales*, Madrid, Aguilar, 1979, vol. 4, p. 404.

diferentes. La teoría de la probabilidad tuvo su origen en el juego de la ruleta; la recopilación de hechos estadísticos nació con la necesidad estatal de soldados y dinero; los seguros marítimos empezaron con los naufragios y la piratería del antiguo Mediterráneo; los modernos estudios de mortalidad tienen sus raíces en los efectos de las pestes del siglo XVII; la teoría de los errores se creó en la astronomía; la teoría de la correlación, en la agricultura; la teoría de las series temporales, en la economía y la meteorología; las teorías del análisis y ordenación de elementos, en la psicología; y la teoría de los métodos de las ji-cuadrado, en la sociología. En retrospectiva, parece como si todas las fases de la vida humana y todas las ciencias hubieran contribuido de alguna forma importante al tema. Escribir su historia es, por tanto, una tarea interesantísima, pero también dificilísima.

Para fijar el origen de la Estadística como ciencia, sin perdernos entre censos chinos y astrágalos egipcios, es conveniente considerar lo que dice al respecto D. PEÑA[18]:

La Estadística actual es el resultado de la unión de dos disciplinas que evolucionan independientemente hasta confluir en el siglo XIX: la primera es el cálculo de probabilidades, que nace en el siglo XVII como teoría matemática de los juegos de azar; la segunda es la «Estadística» (o ciencia del Estado, del latín Status) que estudia la descripción de datos, y tiene unas raíces más antiguas. La integración de ambas líneas de pensamiento de lugar a una ciencia que estudia cómo obtener conclusiones de la investigación empírica mediante el uso de modelos matemáticos.

Aun así, haremos unas breves referencias a la prehistoria de esta ciencia tanto de sus orígenes «estadísticos» como «probabilísticos» .

Las primeras sociedades humanas organizadas han sentido la necesidad de contar con datos numéricos de la población y de sus condiciones materiales de existencia. Esas necesidades nacían fundamentalmente de intereses bélicos y tributarios. Quienes ostentaban el poder querían saber con cuántos hombres y pertrechos contaban para hacer frente a un posible enemigo y cuánto trigo cosechaban o cuántas riquezas atesoraban sus súbditos para que contribuyesen en correspondencia al mantenimiento de la sociedad y de sus jerarcas. Vestigios de estos rudimentarios censos los encontramos en China (Confucio da noticias de estadísticas agrarias y comerciales durante el mandato del emperador Yao en el año 2238 a. de C.), Egipto (el Pentateuco hace referencia a los censos realizados por Moisés), Grecia y Roma. Con la caída del Imperio romano y el período de recesión que se vive en Occidente, la práctica censal desaparece y no vuelve a aparecer hasta el siglo XV con la creación de los Estados modernos.

18. *Op. cit.*, vol. 1, p. 23.

En el siglo XVII se empiezan a perfilar algunos de los conceptos estadísticos básicos, fundamentalmente en lo que se refiere al estudio de datos políticos y económicos y de poblaciones. Por un lado, se constituye en Alemania la llamada escuela descriptiva-histórica cuya finalidad es la descripción de las principales magnitudes concernientes al Estado. Su fundador es H. CORNING (1606-1681), siendo continuada su labor, un siglo después, por G. ACHENWALL (1719-1772), padre del término «estadística». Por otro lado, en Inglaterra, se forma la llamada escuela de aritmética política, que va a tener más interés en el posterior desarrollo de la Estadística y de la Demografía. Uno de sus fundadores, J. GRAUNT (1620-1674), llevó a cabo una estimación de la población inglesa, estudiando la regularidad de ciertas características (relación entre nacimientos masculinos y femeninos, tasas de mortalidad por edades, etc.). Estos trabajos fueron completados por W. PETTY (1623-1687). En esa misma línea E. HALLEY (1656-1742) publicó las primeras tablas de mortalidad, que son el arranque de toda la Estadística Actuarial. Todos estos trabajos se realizan a la sombra de las ideas de la Ilustración, cuyos planteamientos sociales y modernizadores promueven en todos los países europeos los primeros censos y las creaciones de institutos, agencias y asociaciones estadísticas (en España, el primer censo promovido por el conde de Floridablanca data de 1787, pero hay que esperar casi un siglo, 1856, para que se constituya la primera Comisión Nacional de Estadística).

Para buscar los orígenes del Cálculo de Probabilidades no hay que remontarse tantos siglos. A pesar de que el juego es una actividad humana que se practica desde los albores de la humanidad (los trabajos arqueológicos sobre cualquier tiempo y sociedad arrojan innumerables pruebas al respecto con múltiples hallazgos de dados y tabas), hasta el siglo XVI no se tiene noticia de un tratamiento *científico* de los juegos de azar. El motivo de este tardío intento de *dominar* el azar puede encontrarse en el carácter divino que se le atribuye hasta la época renacentista (pensemos que aún hoy la gente acude a brujos y pitonisas para que le muestren su futuro escondido tras el azar de una secuencia de naipes). Las sociedades primitivas explicaban el mundo mediante la voluntad divina. Cuando comprendieron que el viento no era el soplido de Eolo y que la buena cosecha no se debía a sus rogativas a Ceres, se pusieron en la tarea de explicar el mundo sin intervenciones divinas. Pero parece que dejaron un resquicio para matar el aburrimiento de los dioses: el azar. KENDALL[19] va más allá al afirmar que *el cálculo de probabilidad no se hace posible hasta que Europa se libera del dogmatismo de la teología medieval.*

Bien es verdad que mucho antes del siglo XVI filósofos como ARISTÓTELES (384-322 a. de C.) han recogido reflexiones sobre el azar que, en lo fundamental, siguen hoy abiertas. En el capítulo 4, libro segundo, de su *Física*, considera ARISTÓTELES tres actitudes ante el azar. En primer lugar, hay quien considera que su existencia está ligada a nuestra ignorancia, punto de vista mantenido muchos siglos después por P. S. DE LAPLACE (1749-1827) o H. POINCARÉ (1854-1912). En segundo lugar, los que no sólo creen en la existencia del azar, sino que lo consideran el motor inteligente del universo (son muchos hoy en día los especialistas en Biología molecular que señalan el origen aleatorio de las más perfectas formas de vida).

19. *Op. cit.*, p. 406.

Por último, los que piensan que el azar es una causa escondida al razonamiento humano, fruto de una inteligencia divina y sobrenatural.

Fuera de estos planteamientos filosóficos, las primeras consideraciones analíticas del azar tienen su origen en la práctica de los juegos de azar y el interés de los jugadores en conocer sus posibilidades y maximizar sus ganancias. El primer tratado conocido se debe al italiano G. CARDANO (1501-1576) y lleva por título *De ludo alae*. Este autor y el mismo GALILEO (1564-1642) se plantean diversas preguntas sobre el juego de dados, comparando combinaciones más o menos ventajosas.

Pero casi todos los historiadores del Cálculo de Probabilidades coinciden en situar su verdadero origen en la correspondencia mantenida entre B. PASCAL (1623-1662) y P. DE FERMAT (1601-1665) a propósito de ciertas cuestiones sobre el juego planteadas al primero por un empedernido jugador francés, el caballero de MÉRÉ. El problema básico que se les propone es cómo repartir el dinero apostado cuando la partida es interrumpida; como se ve, dicho problema que PASCAL y FERMAT resuelven valiéndose del análisis combinatorio, plantea la medición de la ganancia esperada por un jugador.

Como colofón a los trabajos de los autores citados y de otros como L. PACCIOLI (1445-1514) o N. TARTAGLIA (1499-1557) aparece en 1713 la obra póstuma de J. BERNOULLI (1654-1705). Su *Ars conjectandi*, generalización de una memoria de CH. HUYGENS (1629-1695), *De ratiociniis in ludo aleae*, sistematiza el análisis combinatorio, recoge todas las ideas anteriores sobre los juegos de azar e introduce conceptos nuevos y tan importantes como la convergencia en probabilidad o la ley de los grandes números, a la que él se refería, y no sin razón, como *teorema de oro*.

Cinco años después de la publicación de la obra de J. BERNOULLI, A. DE MOIVRE (1667-1754) publica el tratado *Doctrine of chances*, donde formaliza la ley de los grandes números, obtiene la ley normal como límite de la binomial y demuestra un primer teorema de límite central restringido a variables dicotómicas. Los desarrollos de DE MOIVRE, como los de sus antecesores, siguen teniendo un carácter de *matemática recreativa* de cara a los juegos de azar, pero ya vislumbran algunas aplicaciones actuariales.

El reverendo T. BAYES (1702-1761), también en obra póstuma, estudió las probabilidades de las causas ante un efecto observado, sentando las bases de la interpretación subjetivista de la probabilidad y del enfoque bayesiano de la Estadística.

Llegados a este punto de la historia, empieza a sentirse una doble atracción: por un lado, los «estadísticos» sienten la necesidad de introducir las ideas de probabilidad en sus estudios demográficos para poder dar un salto hacia la modelización; por otro lado, los «probabilistas» han tocado techo en sus juegos, nunca mejor dicho, de salón, y ven que han desarrollado un instrumento demasiado valioso como para malgastarlo en timbas y tugurios. En el ambiente ilustrado de la Francia de finales del XVIII y principios del XIX se empieza a gestar lo que hoy entendemos como Estadística.

P. S. DE LAPLACE (1749-1827) expone en su *Teoría analítica de las probabilidades* las ventajas que ofrece la Estadística para estudiar fenómenos complejos, aplicando las teorías probabilísticas a la mecánica celeste. Por su parte K. F. GAUSS (1777-1855) generaliza la teoría laplaciana de los errores, aplicándola a las

mediciones en astronomía y navegación. Ambos, junto con S. D. POISSON (1781-1840) y el marqués de CONDORCET (1743-1794), aplicaron las ideas probabilísticas a problemas prácticos, influyendo de manera determinante en A. QUETELET (1796-1874), que fue el primero que intuyó las posibilidades que tenía la nueva ciencia para ser utilizada en el estudio de la realidad social. Con su sentencia *la urna a la que interrogamos es la Naturaleza*, cerraba definitivamente el Cálculo de Probabilidades a naipes, dados y bolas (aunque la sencillez de estos problemas recomiende su utilización en el aprendizaje), abriéndolo a la realidad.

Después de este esplendor que se vive en Francia es Gran Bretaña quien tomará la antorcha en esta evolución de la Estadística. Volviendo de nuevo a KENDALL[20]:

> *Si tenemos que elegir una fecha como origen de la moderna teoría de la estadística, podemos señalarla, algo arbitrariamente, en 1890. Francis Galton tenía entonces 68 años, pero todavía tenía por delante 20 años de vida productiva. Un profesor de economía, Francis Ysidro Edgeworth (entonces de 45 años), estaba llamando la atención sobre las regularidades estadísticas en los resultados de las elecciones, en los versos griegos y en el emparejamiento de las abejas, e iba a proponer una notable generalización de la ley del error. Un joven llamado Karl Pearson (de 35 años) se acababa de unir al biólogo Walter Weldon, del University College de Londres y estaba meditando las lecciones que finalmente se convertirían en* The Grammar of Science. *Un estudiante llamado George Udny Yule, a los 20 años de edad, había atraído la atención de Pearson. Y en aquel año nació el más grande de todos ellos: Ronald Aylmer Fisher. En los siguientes cuarenta años, a pesar de las obras rusas sobre teoría de la probabilidad (concretamente las obras de Andrei Markov y Aleksandr Chuprov), el desarrollo de la estadística teórica fue debido sobre todo a autores ingleses. En aquel momento se produjo algo parecido a una explosión intelectual en Estados Unidos y en la India. Francia estaba siguiendo una línea individualista en la teoría de la probabilidad bajo la inspiración de Emile Borel y Paul Levy; e Italia, bajo la influencia de Corrado Gini, se estaba desarrollando también independientemente. Pero al final de la segunda guerra mundial el tema había cruzado todas las fronteras nacionales y se había convertido en una de las disciplinas reconocidas del mundo industrial, tecnológico y científico.*

Bajo el influjo de los trabajos de C. R. DARWIN (1809-1882), F. GALTON (1822-1911) y, posteriormente, K. PEARSON (1857-1936) y R. A. FISHER (1890-1962), ponen las bases de la Estadística moderna. Los estudios del primero sobre la herencia genética inauguran todo un campo de aplicación de la Estadística: las ciencias de la vida. Esta corriente se desarrolla alrededor de la revista *Biometrika*, fundada por PEARSON en 1901. Sus aportaciones teóricas son incontables: regresión y correlación, test de hipótesis, diseño de experimento, estimación, etc.

20. *Op. cit.*, pp. 406 y 407.

Por otro lado, a finales del siglo XIX la Estadística es reconocida en todo su valor por la comunidad científica que ve cómo la ciencia experimental y determinista por antonomasia, la Física, tiene que utilizar los conceptos y métodos estadísticos para enfrentarse a la realidad a escala atómica (mecánica estadística, teoría cinética de los gases, principio de incertidumbre de Heisenberg, etc.).

Intentar un breve resumen de la evolución de la Estadística en el último medio siglo es tarea inútil. El sinfín de nuevos métodos y teorías y el inmenso campo de aplicación dificultan una visión general de lo acontecido y hace arriesgado resaltar alguna innovación decisiva. Pero si hubiera que destacar algunos hechos éstos serían:

- La axiomatización propuesta en 1933 por A. KOLMOGOROV (n. en 1903) partiendo del concepto matemático de medida.
- El espectacular desarrollo de las matemáticas, que proporciona nuevos recursos formales a la Estadística. La teoría de conjuntos y la teoría de la integral de E. BOREL (1871-1956) y H. LEBESGUE (1875-1941), principalmente, junto a la axiomatización de KOLMOGOROV, añaden rigor y universalidad a la Estadística.
- Los nuevos campos de aplicación, principalmente la Economía, que no sólo abren perspectivas a la Estadística, sino que generan ramas científicas nuevas (la Econometría).
- La generalización masiva del uso de los ordenadores y de la informática que introduce la Estadística en problemas prácticos hasta ahora inabordables por imposibilidad de cálculo, a la vez que impulsa teorías apropiadas para el análisis de grandes colecciones de datos.

1.5 Estadística y Ciencias Sociales

Actualmente son muy pocos los campos científicos y las actividades humanas ajenos a la utilización de esta potente herramienta que es la Estadística. Desde sus principios, la Estadística ha mostrado una clara vocación de apego a la realidad, siendo esta vocación su principal motor de desarrollo. Día a día siguen apareciendo nuevas parcelas del saber y de la vida que necesitan del apoyo estadístico. En muchos casos, esas necesidades van por delante de los avances de esta ciencia que tiene que improvisar continuamente nuevos métodos y técnicas demandadas desde los campos más diversos. Sin querer, ni poder, ser exhaustivos, vamos a dar una relación de las ciencias y de las actividades que recurren a la Estadística para encontrar solución a algunos de sus problemas:

- **Economía**. El carácter abrumadoramente cuantitativo de esta disciplina obliga a la utilización de métodos estadísticos que pongan orden en el caos numérico y ayuden a sacar conclusiones. La vitalidad de este maridaje entre Estadística y Economía ha hecho surgir un campo científico nuevo: la Econometría.

- **Demografía**. En el siglo XVIII, con los estudios de J. GRAUNT, W. PETTY y E. HALLEY, sobre tasas de nacimiento y mortalidad, las dos ciencias aún iban de la mano.
- **Psicología y Sociología**. Las técnicas estadísticas tratan de encontrar regularidades en los comportamientos humanos de por sí diversos.
- **Industria**. El control de calidad es hoy en día una pieza indispensable en cualquier proceso productivo.
- **Empresa**. La Estadística proporciona un resumen eficaz de la realidad que rodea al mundo de la empresa y ayuda a tomar decisiones en las lógicas situaciones de incertidumbre. El análisis multivariante es, por otro lado, un arma fundamental en cualquier análisis de mercado.
- **Física**. La imposibilidad de mediciones exactas en la Física de las partículas subatómicas obliga a mediciones probabilísticas. En estos casos la Estadística es más que una simple herramienta, constituyéndose en parte fundamental de muchas de estas teorías: mecánica estadística, movimiento browniano, radiactividad, principio de incertidumbre de Heisenberg, etc.
- **Química**. La Estadística se utiliza tanto en campos teóricos (termodinámica, teoría cinética de los gases, etc.) como prácticos (probar la eficacia de medicamentos en la industria farmacéutica, medir rendimientos de fertilizantes, etcétera).
- **Astronomía**. Desde los primeros balbuceos de la Estadística (Gauss, Laplace, etc.), la Astronomía no ha dejado de utilizarla tanto para controlar los errores en las mediciones como para realizar un «mapa probabilístico» del universo en cada instante.
- **Meteorología**. Las predicciones del tiempo (temperaturas, lluvias, vientos, etc.), necesarias para múltiples actividades, tienen su base científica en el estudio de las series temporales.
- **Genética**. A principios de este siglo los trabajos de K. PEARSON y R. A. FISHER confirmaron las leyes enunciadas por G. MENDEL (1822-1884) que en principio habían pasado desapercibidas para la comunidad científica, tal vez por el *extraño laboratorio* en que se gestaron: el jardín del monasterio donde vivía MENDEL.
- **Biología y Medicina**. La influencia de PEARSON y su *Biometrika* cambiaron radicalmente los hábitos de investigación en estos campos de las ciencias de la vida.
- **Agronomía**. No olvidemos que gran parte del trabajo de R. A. FISHER, sin duda alguna el padre de la Estadística moderna, se desarrolló en un instituto de investigación agrícola, el Rothahamsted Experimental Station, de Harpenden.
- **Lingüística**. Ésta es una ciencia a la que recientemente se ha incorporado la Estadística como un útil más: análisis de textos, comparación de lenguas a fin de encontrar raíces comunes, etc.

- **Paleontología y Arqueología.** Las técnicas de análisis multivariante son una buena herramienta para clasificar, por ejemplo, trozos de vasijas o dientes de mamut.

Otras aplicaciones de la Estadística pueden ir desde la organización del tráfico en una gran ciudad o el control de vuelos en un aeropuerto, hasta la realización de un mapa comarcal de un cierto país o la obtención de una regla de decisión para que un banco nos conceda o deniegue un crédito.

En el ámbito de las Ciencias Sociales es donde la Estadística despliega sus mayores posibilidades. Existen una serie de motivos que hacen necesaria, e incluso obligatoria, esta utilización:

1) La complejidad de los fenómenos objeto de estudio de las Ciencias Sociales.

Los fenómenos sociales se producen en un mundo en constante cambio donde las influencias de otros espacios y otro tiempo son innumerables. Existe la imposibilidad de aislar influencias, causas y efectos, y la Estadística puede servir de alguna forma a discernir entre lo accesorio y lo fundamental, para poner orden en lo diverso.

2) Las Ciencias Sociales no admiten laboratorios en los que poder repetir innumerables veces fenómenos aislados.

El único laboratorio es la propia realidad, siempre cambiante y distinta, donde los fenómenos se ven mediatizados por todo tipo de influencias.

3) El factor humano es determinante en todo fenómeno social.

El hombre, cada hombre, con sus particularidades, ideas, gustos y comportamientos es el sujeto activo de toda realidad social. Podemos prever un comportamiento general, pero nunca el de un solo individuo, y sus decisiones siempre generarán una cadena de circunstancias que redundarán en ese comportamiento general. El individuo, por naturaleza, es en sí imprevisible.

4) El estudio de la realidad social da lugar a inmensas colecciones de datos.

La Estadística pondrá orden en esas observaciones, ayudará a la comprensión de los datos resumiéndolos, sacará a la luz información latente y despreciará la errónea, redundante o inútil. Con la mejora cuantitativa y cualitativa de las bases de datos sobre la realidad social, hoy en día cualquier análisis en este sentido no puede obviar la importancia de la información recogida.

5) El carácter ilimitado, en el tiempo, el espacio y las personas, de los hechos sociales obliga a trabajar con una muestra representativa.

Las características que debe cumplir la selección de la muestra y la recogida de datos para que esta muestra sea representativa del total, y las condiciones y alcance de las conclusiones inferidas a la población, son la finalidad de la Teoría de Muestras.

Las Ciencias Sociales tienen por objeto el estudio del comportamiento del hombre desde diversos prismas. La Estadística servirá para evitar lo particular y recalcar lo general, es decir, para aislar los hechos sociales de aquellos comportamientos particulares. En este sentido se refiere E. DURKHEIM cuando habla de la importancia del empleo de la Estadística en el estudio de la realidad social: *La Estadística expresa cierto estado del alma colectiva.*

Detengámonos algo más en el análisis de las relaciones entre la Estadística y la Economía. Desde los años cincuenta, la Economía se ha convertido en una ciencia positiva, que se desmarca de los elementos normativos e ideológicos como sustento final de la validez de las teorías. Ello implica un alto grado de contrastación, tanto de los supuestos sobre los que descansan las teorías como de las condiciones del entorno, esto es, del contexto en el que son aplicables.

El Cálculo de Probabilidades y la Estadística son asimismo útiles apreciables en el terreno de la gestión. En primer lugar, las técnicas de reducción de datos, de predicción, de muestreo, permiten dibujar una imagen (un resumen) de los escenarios presentes y futuros en los que se actúa. En segundo lugar, suministra a los encargados de la gestión útiles para la toma de decisiones en ambiente de incertidumbre, con la adecuada evaluación de riesgos.

La necesidad e importancia del uso de la Estadística como método de investigación social no debe hacernos caer en tres errores muy comunes:

- La Estadística no es un fin, sino un medio, para el estudioso y el científico de lo social. Esto no hay que olvidarlo a la hora de plantearnos su enfoque y su enseñanza.
- La Estadística no es la *única* técnica de investigación social. En las últimas décadas cierto fantasma cuantitativo recorre los campos científicos invalidando todo estudio en el que los datos numéricos y su análisis no tengan un papel preponderante.
- La Estadística es una ayuda y nunca un sustituto para la formulación de una teoría. Sólo con *estadística* no se va a ninguna parte. Ya lo advertía A. LANG: *Utiliza la Estadística como el borracho las farolas: como apoyo, más que como iluminación.* Unas ideas, una teoría, unas hipótesis se podrán validar con la ayuda estadística, mediante métodos objetivos, que incluso fijarán las delimitaciones de las mismas. Ahora bien, sólo a partir de datos y Estadística nunca podremos *inventar* teorías.

En la aplicación estadística a cualquier campo es necesario que exista una estrecha relación entre el estadístico y el especialista; aún mejor, es conveniente que el estadístico tenga unas ciertas ideas de la realidad donde se va a aplicar su instrumento y que el especialista posea unas nociones de los métodos estadísticos y sus posibilidades, para que no espere más de lo que éstos le pueden proporcionar y sepa plantear el problema en los justos términos en los que la Estadística puede abordarlo.

Capítulo 2

CONCEPTO DE PROBABILIDAD

2.1 Fenómenos aleatorios

En Estadística se emplea la palabra **experimento** para designar todo acto que proporciona unos datos. Así, llamamos experimento a una operación destinada a provocar un fenómeno (lanzamiento de un dado, bombardeo de un átomo con neutrones, estiramiento de un cable hasta su ruptura, etc.), pero, también, a una observación de un fenómeno real (medir la altura de un individuo, contabilizar la renta de una familia, apuntar el número de bajas diarias en una empresa, etc.).

En algunas situaciones, la realización sucesiva de un experimento en las mismas condiciones produce el mismo resultado: son los llamados **experimentos deterministas**. A los fenómenos que se manifiestan en dichos experimentos los llamaremos **fenómenos deterministas**. Son fenómenos deterministas la caída libre de los cuerpos (ley de la gravedad) o la formación de agua a partir de oxígeno e hidrógeno bajo ciertas condiciones.

Pero también hay situaciones en las que la realización sucesiva de un experimento en las mismas condiciones produce resultados distintos: son los llamados **experimentos aleatorios**. A los fenómenos que se manifiestan en dichos experimentos los llamaremos **fenómenos aleatorios**. Son fenómenos aleatorios el lanzamiento de un dado, el rendimiento de una semilla, la duración de la vida de una persona, el alcance de un proyectil, el número de kilovatios que consume cada día una ciudad, etc. Algunos autores no incluyen entre estos últimos los llamados experimentos casuales, cuyo comportamiento es errático, es decir, no obedecen a ninguna pauta.

En los experimentos deterministas las mismas causas producen los mismos efectos, mientras que en los aleatorios las mismas causas producen distintos efectos. Por ello, los resultados de un experimento determinista se pueden predecir, no así los de un experimento aleatorio.

Aunque la definición que daremos a continuación no es muy precisa, ayuda a enmarcar los experimentos aleatorios, diferenciándolos de los deterministas y de los casuales.

Definición 2.1 *Decimos que un experimento es aleatorio si verifica las siguientes condiciones:*

1. *Antes de realizar el experimento se conocen todos sus posibles resultados.*
2. *El resultado concreto de una realización del experimento no es predecible (este hecho los distingue de los experimentos deterministas).*
3. *Posee la propiedad denominada de* regularidad estadística, *esto es, un comportamiento predecible a largo plazo (esta propiedad los distingue de los experimentos casuales).*

OBSERVACIONES

1) Por comodidad, una vez vistas sus diferencias conceptuales, utilizaremos *fenómeno* y *experimento* como sinónimos.

2) En la práctica, la mayoría de los experimentos de interés no son repetibles en las mismas condiciones. Los resultados estarán influidos por múltiples factores no controlables. Si estudiamos la duración de la vida de un hombre, no podremos repetir el experimento bajo las mismas condiciones; a lo sumo dichas condiciones serán similares. De todas formas, esta consideración no afecta globalmente al concepto de experimento y fenómeno aleatorio.

3) La distinción entre fenómeno determinista y aleatorio es meramente metodológica. De hecho, desde el punto de vista del Cálculo de Probabilidades, los primeros se pueden considerar como un caso particular de los segundos.

Como señalamos en la definición anterior, al realizar un experimento aleatorio llevamos a cabo una operación al final de la cual obtendremos un resultado; éste es *a priori* imprevisible, pero pertenece a un conjunto que puede ser descrito completamente de antemano.

Definición 2.2 *Dado un experimento aleatorio, llamamos espacio muestral al conjunto de todos los resultados posibles de un experimento.*

Al espacio muestral lo denotaremos por Ω y a sus elementos, que llamaremos puntos muestrales, por ω.

A todo fenómeno o experimento aleatorio le asociaremos un espacio muestral.

$$\boxed{\text{EXPERIMENTO ALEATORIO}} \longrightarrow \boxed{\text{ESPACIO MUESTRAL}}$$

Por ejemplo, si lanzamos un dado y observamos el resultado obtenido, el espacio muestral asociado será:

$$\boxed{\text{Lanzamiento de un dado}} \longrightarrow \boxed{\Omega = \{1, 2, 3, 4, 5, 6\}}$$

Al asignar a un experimento aleatorio un espacio muestral hacemos una idealización o simplificación, prescindiendo de resultados raros (por ejemplo, que el dado caiga apoyado en una arista o en un vértice, o que no caiga).

Llamamos **suceso** a todo subconjunto A del espacio muestral ($A \subset \Omega$). Por ejemplo, en el espacio muestral anterior es un suceso *obtener un as*, $\{1\}$, u *obtener par*, $\{2,4,6\}$.

El conjunto de todos los sucesos es, por tanto, el conjunto de **partes de** Ω, $\mathcal{P}(\Omega)$. De entre todos esos sucesos, algunos sólo contendrán un punto muestral (**sucesos elementales**), mientras que otros contendrán más de un punto (**sucesos compuestos**). El suceso *obtener un seis*, $\{6\}$, es elemental; en cambio *obtener impar*, $\{1,3,5\}$, es un suceso compuesto. Cuando realizamos un experimento el resultado observado es un punto muestral; afirmaremos, entonces, que ha ocurrido un suceso cualquiera cuando ocurre un punto muestral contenido en el mismo. Así, si al lanzar un dado obtenemos un dos, podemos decir que han ocurrido, entre otros, los sucesos $\{2\}$ y $\{2,4,6\}$. Por tanto, diremos que ha ocurrido el suceso A si ha ocurrido un ω perteneciente a A.

De entre todos los sucesos existen dos especialmente singulares: el **suceso imposible** (aquel que nunca ocurre), $\emptyset$, y el **suceso seguro** (aquel que ocurre siempre), Ω.

En el experimento de lanzar un dado, el conjunto de las partes, es decir, el conjunto de todos los sucesos es:

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \ldots, \{6\}, \{1,2\}, \ldots, \{1,2,3\}, \ldots, \Omega\}$$

donde los sucesos elementales son $\{1\}, \ldots, \{6\}$; el resto son compuestos.

En definitiva, hemos establecido una correspondencia entre sucesos y conjuntos, lo cual nos va a permitir trabajar con sucesos como lo hacíamos con los subconjuntos de un conjunto.

2.2 Operaciones y relaciones entre sucesos. Concepto de σ-álgebra

Establecida ya la relación entre conjuntos y sucesos, recordemos algunas operaciones y relaciones entre conjuntos, que ahora serán de interés para trasladarlas a los sucesos.

2.2.1 Operaciones entre sucesos

1) Complementario de un suceso. Sea A un suceso de un espacio muestral Ω. Llamaremos suceso complementario de A al formado por los puntos muestrales que no pertenecen a A, es decir:

$$\overline{A} = \{\omega \in \Omega \mid \omega \notin A\}\,.$$

Obviamente, si ocurre $\overline{A}$ no ocurre A y viceversa. En concreto, $\overline{\overline{A}} = A$, como se comprueba fácilmente. Por este motivo, al suceso complementario se le denomina también contrario.

El diagrama de VENN que se muestra en la figura 2.1 ilustra esta definición.

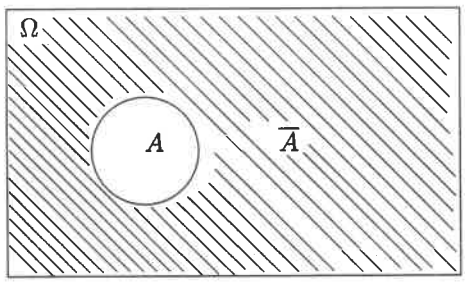


FIG. 2.1.

Por ejemplo, en el experimento de lanzar un dado, el suceso complementario de *obtener par* es *obtener impar*.

2) Unión de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Se define la unión de A y B como el suceso formado por todos los puntos muestrales que pertenecen al menos a uno de los dos sucesos, es decir:

$$A \cup B = \{\omega \in \Omega \mid \omega \in A \quad \text{u} \quad \omega \in B\} .$$

Si ocurre $A \cup B$, ocurre A, B o ambos. Por este motivo, el suceso $A \cup B$ se denomina también A o B. Entre otras, la unión de sucesos cumple las propiedades conmutativa y asociativa.

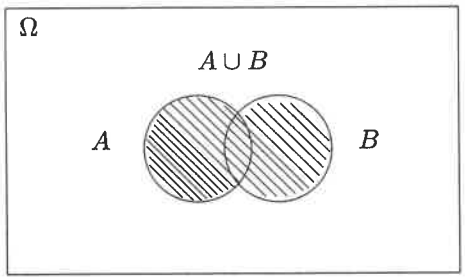


FIG. 2.2.

Por ejemplo, la unión de los sucesos *obtener par*, $\{2,4,6\}$, y *obtener número primo*, $\{1,2,3,5\}$ es el suceso *obtener par o primo*, $\{1,2,3,4,5,6\}$, es decir, Ω.

3) Intersección de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Se define la intersección de A y B como el suceso formado por todos los puntos muestrales que pertenecen a ambos, es decir:

$$A \cap B = \{\omega \in \Omega \mid \omega \in A \quad \text{y} \quad \omega \in B\} .$$

Si ocurre $A \cap B$, ocurren A y B. El suceso $A \cap B$ se denomina también producto de A y B y, si no se presta a confusión, simplificaremos la notación, escribiendo AB. Entre otras, la intersección de sucesos cumple las propiedades conmutativa y asociativa. Además, la unión y la intersección de sucesos verifican la propiedad distributiva.

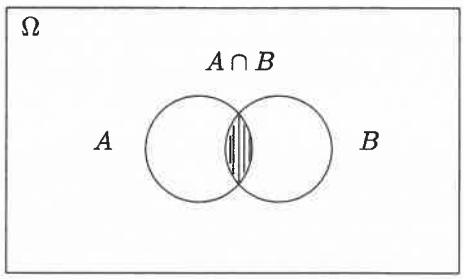


FIG. 2.3.

Por ejemplo, la intersección de los sucesos *obtener par*, $\{2,4,6\}$, y *obtener número primo*, $\{1,2,3,5\}$, es el suceso *obtener par y primo*, $\{2\}$.

Las definiciones de unión e intersección se generalizan para colecciones finitas o infinitas de sucesos.

Las leyes de MORGAN relacionan las tres operaciones anteriores:

1) $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

2) $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

La demostración de estas dos igualdades no presenta dificultad, pudiéndose generalizar los resultados a una colección finita o infinita de sucesos.

4) Diferencia de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Se define la diferencia de A y B como el suceso formado por todos los puntos muestrales que pertenecen a A y no pertenecen a B, es decir:

$$A - B = \{\omega \in \Omega \mid \omega \in A \quad \text{y} \quad \omega \notin B\} .$$

Si ocurre $A - B$, ocurre A pero no ocurre B. Es evidente que

$$A - B = A \cap \overline{B} .$$

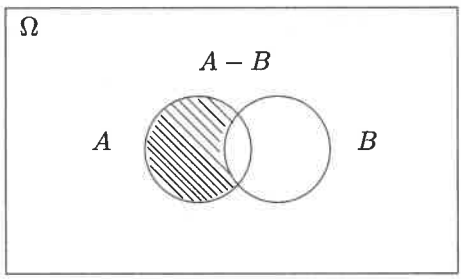


FIG. 2.4.

Por ejemplo, la diferencia entre el suceso *obtener par*, $\{2,4,6\}$, y el suceso *obtener número primo*, $\{1,2,3,5\}$, esto es, $\{2,4,6\} - \{1,2,3,5\}$, resulta el suceso $\{4,6\}$.

5) Diferencia simétrica de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Se define la diferencia simétrica de A y B como el suceso formado por todos los puntos muestrales que pertenecen a A o a B, pero no a ambos, es decir:

$$A\Delta B = \{\omega \in \Omega \mid (\omega \in A \quad \text{u} \quad \omega \in B) \quad \text{y} \quad \omega \notin A \cap B\} .$$

Si ocurre $A\Delta B$, ocurren A o B, pero no los dos a la vez. Por definición:

$$A\Delta B = (A \cup B) - (A \cap B) .$$

Se comprueba que la diferencia simétrica puede calcularse como

$$A\Delta B = (A - B) \cup (B - A) .$$

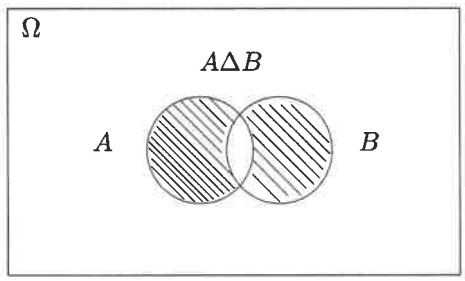


FIG. 2.5.

Por ejemplo, la diferencia simétrica de los sucesos *obtener par*, $\{2,4,6\}$, y *obtener número primo*, $\{1,2,3,5\}$, es el suceso $\{1,3,4,5,6\}$.

2.2.2 Relaciones entre sucesos

1) Inclusión de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Diremos que el suceso A está contenido o incluido en el suceso B, si todos los puntos muestrales de A pertenecen a B, es decir:

$$\omega \in A \Rightarrow \omega \in B\,.$$

Esta relación la denotaremos como $A \subset B$. Siempre que ocurre A ocurre B.

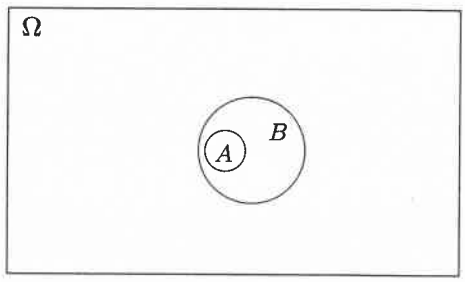


FIG. 2.6.

Por ejemplo, el suceso *obtener impar* está contenido en el suceso *obtener número primo.*

2) Igualdad de sucesos. Sean A y B dos sucesos de un espacio muestral Ω. Diremos que el suceso A es igual al suceso B, si todos los puntos muestrales de cada uno de ellos están en el otro, es decir:

$$\omega \in A \Longleftrightarrow \omega \in B\,.$$

Esta relación la denotaremos como $A = B$. A y B ocurren o no ocurren simultáneamente. Se comprueba que:

$$A = B \Longleftrightarrow A \subset B \quad \text{y} \quad B \subset A\,.$$

Por ejemplo, el suceso *obtener primo y par* es igual o equivalente al suceso *obtener un dos.*

3) Sucesos incompatibles, disjuntos o mutuamente excluyentes. Sean A y B dos sucesos de un espacio muestral Ω. Diremos que el suceso A es incompatible con B, si ningún punto muestral de Ω es común a ambos, es decir:

$$A \cap B = \emptyset\,.$$

La ocurrencia de A impide la de B y viceversa. Se dice también que A y B son sucesos disjuntos o mutuamente excluyentes.

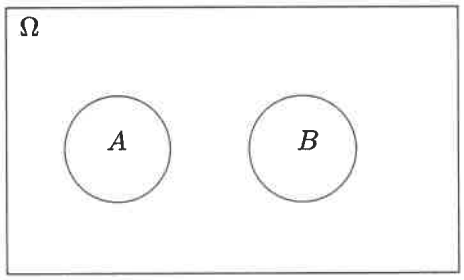


FIG. 2.7.

Por ejemplo, el suceso *obtener par* y el suceso *obtener un cinco* son incompatibles.

2.2.3 Concepto de σ-álgebra

En muchas ocasiones no nos interesarán todos los sucesos de un experimento (todos los subconjuntos de Ω), sino únicamente una parte de ellos (ciertos subconjuntos de $\mathcal{P}(\Omega)$). Un ejemplo ilustrará la conveniencia de considerar solamente algunos sucesos. Dos jugadores lanzan un dado y apuestan a *pares* o *impares*. Es evidente que en esta situación no interesarán sucesos como *obtener un dos*, $\{2\}$, sino solamente los sucesos *obtener par* y *obtener impar*. Éste es el motivo de introducir un nuevo concepto[1].

Definición 2.3 *Llamaremos σ-álgebra de* BOOLE *a un subconjunto α de $\mathcal{P}(\Omega)$, no vacío, que verifique las siguientes propiedades:*

1. *El complementario de un suceso de α también pertenece a α, es decir,*

$$A \in \alpha \Rightarrow \overline{A} \in \alpha \, .$$

2. *La unión de una colección infinita numerable de sucesos de α también pertenece a α, es decir,*

$$\{A_i\}_{i=1}^{\infty} \quad con \quad A_i \in \alpha \quad \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \alpha \, .$$

Antes justificábamos la necesidad de trabajar con una parte de $\mathcal{P}(\Omega)$, pero, ¿por qué exigimos que dicha parte cumpla las propiedades anteriores y no otras?

1. La razón alegada para introducir el concepto de σ-álgebra es mucho más profunda que lo que da a entender la ingenuidad del ejemplo ilustrativo anterior. La necesidad real de su introducción se verá posteriormente, estando ligada a la asignación de probabilidades en espacios muestrales continuos mediante integrales definidas.

En general, existen motivos de coherencia que obligan a que la selección de sucesos que realicemos tenga cierta consistencia interna, y ello queda garantizado con la selección de una σ-álgebra. Por ejemplo, si nos interesa la ocurrencia de un suceso, lógicamente nos va a interesar el que no ocurra dicho suceso (si interesa A, interesa $\overline{A}$); si nos interesan dos sucesos, también el que ocurra uno u otro (si interesan A y B, interesa $A \cup B$), etc. De hecho, de la definición se deducen las siguientes propiedades:

1) La unión de una colección finita de sucesos de α también pertenece a α, es decir,

$$\{A_i\}_{i=1}^n \quad \text{con} \quad A_i \in \alpha \quad \Rightarrow \bigcup_{i=1}^{n} A_i \in \alpha \, .$$

Demostración. Construimos una colección infinita numerable de sucesos de la σ-álgebra mediante la repetición reiterada de los sucesos $A_1, \ldots, A_n$:

$$A_{kn+i} = A_i, \qquad k = 1, 2, \ldots; \quad i = 1, \ldots, n \, .$$

Evidentemente, $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \alpha$. ♣

2) El suceso seguro, Ω, pertenece a α.

Demostración. Como α es no vacío, entonces existe al menos un suceso A en α. Por la propiedad 1 de la definición, también $\overline{A}$ estará en α; entonces Ω estará en α pues $\Omega = A \cup \overline{A}$. ♣

3) El suceso imposible, $\emptyset$, pertenece a α.

Demostración. Es inmediata a partir del resultado anterior, puesto que se tiene que $\emptyset = \overline{\Omega}$. ♣

4) La intersección de una colección infinita numerable de sucesos de α también pertenece a α, es decir,

$$\{A_i\}_{i=1}^{\infty} \quad \text{con} \quad A_i \in \alpha \quad \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \alpha \, .$$

Demostración. Para todo suceso A_i de α se tiene que $\overline{A}_i$ también pertenece a α; luego $\bigcup_{i=1}^{\infty} \overline{A}_i \in \alpha$. Aplicando una de las leyes de Morgan se obtiene:

$$\bigcup_{i=1}^{\infty} \overline{A}_i = \overline{\bigcap_{i=1}^{\infty} A_i} \in \alpha$$

entonces $\bigcap_{i=1}^{\infty} A_i$ pertenecerá a α por pertenecer su complementario. ♣

El mismo resultado se demuestra para una colección finita de sucesos.

En ciertas situaciones será de interés considerar una parte de la σ-álgebra, formada por una colección exhaustiva y excluyente de sucesos de la misma.

Definición 2.4 *Llamamos partición de un conjunto Ω a una colección de subconjuntos de Ω, $\{A_i\}_{i=1}^{\infty}$, tal que*

$$\Omega = \bigcup_{i=1}^{\infty} A_i \ ,$$

siendo dichos conjuntos disjuntos dos a dos, es decir, $A_i \cap A_j = \emptyset$ para cualquier i, j, con $i \neq j$.

Aunque la definición anterior se ha explicitado para una colección infinita numerable, es igualmente válida para colecciones finitas.

Evidentemente, estaremos interesados en particiones de un espacio muestral Ω formadas por sucesos de la σ-álgebra.

Ejemplo 2.1. Realizamos el experimento de lanzar un dado y observar la puntuación obtenida. Supongamos que sólo estamos interesados en el suceso P, *obtener número par*, ¿cuál es la σ-álgebra más pequeña que contiene al suceso P?

Al contener a $P = \{2, 4, 6\}$, esta σ-álgebra, α_1, también contendrá a su complementario, $\overline{P} = I = \{1, 3, 5\}$ (primera condición que tiene que cumplir una σ-álgebra).

Además, α_1 debe incluir necesariamente los sucesos Ω y $\emptyset$ (segunda y tercera propiedad deducidas de la definición de σ-álgebra). Téngase en cuenta, por otra parte, que los sucesos Ω y $\emptyset$ son precisamente la unión y la intersección de P e I.

Por tanto, la menor σ-álgebra que contiene al suceso P es:

$$\alpha_1 = \{\emptyset, P, I, \Omega\} \ .$$

Si estamos interesados en que la σ-álgebra contenga los sucesos *obtener par* y *obtener un uno*, ¿cuál será la σ-álgebra mínima que tenemos que construir?

Podemos seguir el procedimiento anterior para generar la σ-álgebra, α_2, engendrada por los sucesos $\{2, 4, 6\}$ y $\{1\}$, es decir, formar todas las uniones, intersecciones y complementarios. Sin embargo, este método puede ser confuso, con el riesgo de *olvidar* alguno se los sucesos de la σ-álgebra.

Facilitaremos la construcción de la σ-álgebra más pequeña que contenga a unos determinados sucesos considerando los átomos de la misma. Entendemos por **átomos** de una σ-álgebra una colección exhaustiva y excluyente de sucesos de la misma, es decir, una partición de Ω formada por sucesos de la σ-álgebra.

Si estamos interesados en construir la menor σ-álgebra que contenga a unos determinados sucesos, consideraremos en primer lugar estos sucesos y sus complementarios. A continuación, hallaremos las intersecciones de todas las posibles colecciones finitas o infinitas numerables de los mismos. Finalmente, de entre todos los sucesos obtenidos, se seleccionarán aquellos que formen una partición de Ω. Estos sucesos serán los átomos que generarán la σ-álgebra más pequeña deseada.

Los átomos de la σ-álgebra mínima que contiene a los sucesos *obtener par* y *obtener un uno* son:

$$\begin{aligned} A_1 &= \{2, 4, 6\} \ , \\ A_2 &= \{1\} \quad \text{y} \\ A_3 &= \{3, 5\} \ . \end{aligned}$$

Una vez determinados, construiremos los sucesos de la σ-álgebra haciendo todas las uniones posibles de dos, tres o más átomos, añadiendo el suceso imposible. Por tanto, la σ-álgebra buscada será:

$$\alpha_2 = \{\emptyset, \{2,4,6\}, \{1\}, \{3,5\}, \{1,2,4,6\}, \{1,3,5\}, \{2,3,4,5,6\}, \Omega\} .$$

De igual forma construiríamos, por ejemplo, la menor σ-álgebra, α_3, que contuviera a los sucesos *obtener par* y *obtener primo*. En este caso los átomos serían:

$$\begin{array}{rcl} A_1 & = & \{1,3,5\} , \\ A_2 & = & \{4,6\} \quad \text{y} \\ A_3 & = & \{2\} . \end{array}$$

Si estamos interesados en la σ-álgebra que contenga a todos los sucesos elementales, éstos formarían los átomos de la misma. Cualquier subconjunto A de Ω se puede poner como unión de dichos sucesos elementales, es decir, de los átomos; por tanto, la σ-álgebra que buscamos es la más grande, la que contiene todos los subconjuntos de Ω, es decir, $\mathcal{P}(\Omega)$. $\triangleleft$

Algunas σ-álgebras singulares son $\alpha_1 = \mathcal{P}(\Omega)$ y $\alpha_2 = \{\emptyset, \Omega\}$; α_1 contiene todos los sucesos, mientras que α_2 sólo contiene los indispensables para ser una σ-álgebra. En muchas ocasiones usaremos la σ-álgebra $\mathcal{P}(\Omega)$, sobre todo cuando el espacio muestral, Ω, sea finito o infinito numerable.

A los sucesos de la σ-álgebra los denominaremos **sucesos estocásticos**, y, dado que siempre trabajaremos con dichos sucesos, los llamaremos por convenio simplemente **sucesos**.

El par formado por el espacio muestral y una σ-álgebra, (Ω, α), se denomina **espacio probabilizable**.

2.3 Definición de probabilidad

Sea Ω el espacio muestral ligado al experimento aleatorio y sea α una σ-álgebra de sucesos de Ω, es decir, una selección de subconjuntos de Ω. En otras palabras, sea (Ω, α) un espacio probabilizable.

Sea A un suceso; se trata de definir una medida teórica de la ocurrencia de A, es decir, lo que entenderemos por la probabilidad de A.

Históricamente, el concepto de probabilidad se introdujo a partir del de frecuencia, proponiéndose inicialmente una definición hoy en día superada. Expongamos dicha definición, recordando previamente el concepto de **frecuencia relativa** y algunas de sus propiedades.

Realizado el experimento N veces, llamamos **frecuencia absoluta** del suceso A al número de veces que éste se presenta, y lo denotamos por n_A. Denominamos **frecuencia relativa** de dicho suceso al cociente entre su frecuencia absoluta y el número de experimentos, es decir, $fr(A) = n_A/N$. La frecuencia relativa no es más que la medida relativa y empírica de la ocurrencia de un suceso; relativa en cuanto que viene dada por una proporción, y empírica en cuanto que contabiliza hechos ya acaecidos. Veamos algunas de sus propiedades:

1) $0 \leq fr(A) \leq 1$.

Demostración. Evidentemente se tiene $0 \leq n_A \leq N$, con lo que dividiendo por N se obtiene el resultado. ♣

2) $fr(\Omega) = 1$.

Demostración. Como $n_\Omega = N$, su frecuencia relativa es la unidad. ♣

3) $fr(A \cup B) = fr(A) + fr(B)$ si $A \cap B = \emptyset$.

Demostración. $fr(A \cup B) = n_{A \cup B}/N$. Pero como A y B son incompatibles, $n_{A \cup B} = n_A + n_B$; en consecuencia, $fr(A \cup B) = (n_A + n_B)/N = fr(A) + fr(B)$. ♣

La anterior propiedad puede generalizarse a la unión de un número finito o infinito numerable de sucesos.

Si lanzásemos al aire sucesivamente una moneda, en la mayoría de las ocasiones comprobaríamos que, en la medida en que aumente el número de tiradas, la frecuencia relativa del suceso c, *obtener cara*, se irá acercando a $1/2$. Esto constituye un hecho empírico cuya generalización a un experimento cualquiera se conoce como *ley de regularidad estadística* y que podemos enunciar del siguiente modo: *la frecuencia relativa de un suceso se estabiliza cuando el número de tiradas crece indefinidamente.*

Esta ley permite dar una definición frecuentista de probabilidad como el límite de la frecuencia relativa de un suceso cuando el número de tiradas tiende a infinito, es decir,

$$\lim_{N \to \infty} fr(A) = p(A) .$$

Esta definición, hoy en desuso, no es operativa ni congruente con el concepto matemático de límite[2], pero consigue dar una idea intuitiva de lo que se pretende al definir la probabilidad de un suceso. En el capítulo 13 veremos que un concepto análogo al que muestra esta definición no es otra cosa que la *ley de los grandes números.* Daremos a continuación la definición axiomática de probabilidad que introdujo el matemático ruso KOLMOGOROV en 1933; con ella tratamos de dar una medida relativa y teórica de la ocurrencia de un suceso.

Definición 2.5 *Sea un espacio probabilizable* (Ω, α). *Llamamos probabilidad sobre* α *a una aplicación* p, *de* α *en los números reales, que verifica las siguientes propiedades:*

1. $p(A) \geq 0$ *para todo* $A \in \alpha$.

2. No se conoce una expresión general de $fr(A)$ en función de N. Por otro lado, no es posible lanzar una moneda una infinidad de veces. En definitiva, el límite no es calculable analíticamente, y ningún límite se puede calcular experimentalmente.

2. $p(\Omega) = 1$.

3. $p(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} p(A_i)$ *para toda colección infinita numerable de sucesos* A_i *de* α, *disjuntos dos a dos.*

Por dicha aplicación

$$p : \alpha \to \mathbb{R} ,$$

asociamos a todo suceso A un número real, $p(A)$, que es la probabilidad de dicho suceso.

A la terna (Ω, α, p) la llamaremos **espacio probabilístico**.

OBSERVACIONES

1) Nótese que las propiedades de la probabilidad son paralelas a las de la frecuencia relativa. Así, mientras ésta es una medida empírica de la ocurrencia de un suceso, la probabilidad es una medida teórica de la misma; es decir, una medida que trata de evaluar la posibilidad de que ocurra cada suceso antes de realizar el experimento.

2) Cualquier aplicación que verifique la definición anterior es una probabilidad, no teniendo por qué ajustarse a un fenómeno aleatorio real. Evidentemente, ante un determinado fenómeno interesa construir una probabilidad que lo describa lo mejor posible. Cuando la naturaleza del experimento permita suponer una cierta *geometría* en la ocurrencia de los distintos resultados, la asignación de probabilidad se llevará a cabo de forma más o menos mecánica (lanzamiento de dados, elección de cartas en una baraja, selección de individuos en una población, etc.). En caso contrario, la asignación se hará en base a la información histórica que se tenga del experimento en cuestión. El problema de asignar una *buena* probabilidad a un fenómeno aleatorio es el objetivo central de la Inferencia Estadística.

Ejemplo 2.2. Al tirar una moneda, el espacio muestral asociado es $\Omega = \{c, z\}$, siendo c el resultado *obtener cara* y z el resultado *obtener cruz*. Una posible σ-álgebra será $\alpha = \{\emptyset, \{c\}, \{z\}, \Omega\}$ y dos posibles probabilidades, entre muchas, serán:

$$p_1 : \begin{array}{ccc} \alpha & \to & \mathbb{R} \\ \emptyset & \hookrightarrow & 0 \\ \{c\} & \hookrightarrow & 1/3 \\ \{z\} & \hookrightarrow & 2/3 \\ \Omega & \hookrightarrow & 1 \end{array} \qquad p_2 : \begin{array}{ccc} \alpha & \to & \mathbb{R} \\ \emptyset & \hookrightarrow & 0 \\ \{c\} & \hookrightarrow & 1/2 \\ \{z\} & \hookrightarrow & 1/2 \\ \Omega & \hookrightarrow & 1 \end{array}$$

Puede el lector comprobar que, tanto p_1 como p_2, verifican la definición de KOLMOGOROV. Ahora bien, si la moneda es normal, es decir, no está trucada en ningún sentido, es lógico pensar que p_2 es una probabilidad más adecuada para describir el experimento, al asignar idéntica probabilidad a la cara y a la cruz.

Por otra parte, si definimos la siguiente aplicación

$$\begin{array}{rccc} p_3: & \alpha & \to & \mathbb{R} \\ & \emptyset & \hookrightarrow & 0 \\ & \{c\} & \hookrightarrow & 1/4 \\ & \{z\} & \hookrightarrow & 3/5 \\ & \Omega & \hookrightarrow & 1 \end{array}$$

el lector puede comprobar que ésta no verifica la definición de KOLMOGOROV.

En definitiva, p_1 y p_2 son probabilidades y, en cambio, p_3 no lo es. Pero p_1 es una *mala* probabilidad, mientras que p_2 es *buena* en el sentido de que explica bien el comportamiento aleatorio de la moneda. Son éstas, y no otras, las probabilidades que nos interesan. ◁

¿Contradice la definición de KOLMOGOROV la idea común de probabilidad como *cociente entre el número de casos favorables y el número de casos posibles*? Como veremos al final del capítulo, esta idea de probabilidad, introducida por el matemático francés LAPLACE, es un caso particular de la definición de KOLMOGOROV.

2.4 Propiedades de las probabilidades

Sea un espacio probabilístico (Ω, α, p). Entonces:

1) El suceso imposible tiene probabilidad nula, es decir, $p(\emptyset) = 0$.

Demostración. Sea $A_1 = \emptyset, \ldots, A_n = \emptyset, \ldots$ una colección infinita numerable de sucesos. Evidentemente, $\bigcup_{i=1}^{\infty} A_i = \emptyset$ y dichos sucesos son disjuntos dos a dos. Consecuentemente,

$$p(\emptyset) = p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i) = \sum_{i=1}^{\infty} p(\emptyset)\ .$$

Por tanto,

$$p(\emptyset) = 0\ . \clubsuit$$

2) Para toda colección finita de sucesos de α, $\{A_i\}_{i=1}^{n}$, disjuntos dos a dos:

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i)\ .$$

Demostración. Se completa la colección de sucesos $A_1, \ldots, A_n$ con $A_{n+1} = \emptyset$, $A_{n+2} = \emptyset, \ldots$ Con ello, hemos formado una colección infinita numerable de sucesos disjuntos dos a dos que verifican:

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i\ .$$

Sin más que tomar probabilidades en la igualdad anterior,

$$p\left(\bigcup_{i=1}^{n} A_i\right) = p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i) = \sum_{i=1}^{n} p(A_i) + \sum_{i=n+1}^{\infty} p(A_i) = \sum_{i=1}^{n} p(A_i)\ . \clubsuit$$

3) $p(\overline{A}) = 1 - p(A)$ para todo $A \in \alpha$.

Demostración. Dado que $\Omega = A \cup \overline{A}$, $p(A) + p(\overline{A}) = p(\Omega) = 1$. ♣

4) Para todo $A, B \in \alpha$ tal que $A \subset B$, $p(A) \leq p(B)$.

Demostración. Si $A \subset B$ se tiene que $B = A \cup (B - A)$ donde A y $B - A$ son disjuntos. Por tanto, $p(B) = p(A) + p(B - A)$ y, como la probabilidad siempre es no negativa, se deduce el resultado. ♣

Dos consecuencias de la propiedad anterior son las siguientes:

5) $0 \leq p(A) \leq 1$.

Demostración. Dado que $A \subset \Omega$, entonces $p(A) \leq 1$.♣

6) Para todo $A, B \in \alpha$ tal que $A \subset B$, $p(B - A) = p(B) - p(A)$.

Demostración. Está implícita en la de la propiedad 4). ♣

7) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ para todo $A, B \in \alpha$.

Demostración. El suceso $A \cup B$ se puede descomponer como unión de dos sucesos disjuntos,

$$A \cup B = A \cup (B - A) = A \cup (B \cap \overline{A}) .$$

Tomando probabilidades obtenemos:

$$p(A \cup B) = p(A) + p(B \cap \overline{A}) .$$

Por su parte, el conjunto B se puede expresar también como unión de disjuntos,

$$B = B \cap \Omega = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A}) .$$

Tomando de nuevo probabilidades, $p(B) = p(B \cap A) + p(B \cap \overline{A})$; sustituyendo en la expresión de $p(A \cup B)$ obtenemos el resultado. ♣

Las dos siguientes propiedades son consecuencias inmediatas de la anterior:

8) $p(A \cup B) \leq p(A) + p(B)$ para todo $A, B \in \alpha$.

9) Para todo $A, B, C \in \alpha$,

$$p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(AB) - p(AC) - p(BC) + p(ABC) .$$

Demostración.

$$p(A \cup B \cup C) = p[(A \cup B) \cup C] = p(A \cup B) + p(C) - p[(A \cup B) \cap C] .$$

Desarrollando $p(A \cup B)$ de acuerdo con la propiedad 7 y operando con el último término de la expresión,

$$p[(A \cup B) \cap C] = p[(A \cap C) \cup (B \cap C)] = p(A \cap C) + p(B \cap C) - p(A \cap B \cap C)\,,$$

obtendremos el resultado. ♣

Por inducción se generaliza esta propiedad para n sucesos:

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i) - \sum_{\substack{i,j=1 \\ i \neq j}}^{n} p(A_i A_j) + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} p(A_i A_j A_k) + \cdots$$

OBSERVACIONES

1) Todas estas propiedades tienen su correspondencia con propiedades de la frecuencia relativa.

2) Los sucesos $\emptyset$ y Ω no tienen por qué ser los únicos sucesos con probabilidad 0 y 1.

3) Una vez vistas definición y propiedades, la probabilidad resulta ser una medida relativa y teórica de la ocurrencia de un suceso que toma valores entre 0 y 1. Aunque la definición de probabilidad dada por KOLMOGOROV aparenta ser puramente abstracta, intenta formalizar la idea de que una probabilidad es una medida relativa (porcentual) y teórica de la ocurrencia de un suceso. Resultan entonces razonables, tanto los axiomas que la definen, como las propiedades que de ellos se derivan. Por ejemplo, es razonable que un suceso que siempre ocurre tenga probabilidad 1 (axioma 2 de KOLMOGOROV) y que el que nunca ocurre tenga probabilidad cero (propiedad 1); razonable es también que, si siempre que sucede A sucede B, la medida de la ocurrencia de A sea menor o igual que la de B (propiedad 4). De igual manera justificaríamos el resto de axiomas y propiedades.

Una definición axiomática trata de describir idealmente una cierta situación. Este tipo de definiciones *funcionan* en la medida en que ella y los resultados que de ella se deducen sean razonables y coherentes con la realidad que tratan de idealizar. Los comentarios anteriores nos permiten afirmar que la definición de KOLMOGOROV es una buena definición.

Ejemplo 2.3. Un estudio sobre la situación sociolaboral de las familias de una población ha arrojado los siguientes resultados:

- En el 70 % de los hogares trabaja al menos el marido.
- En el 20 % al menos trabaja la mujer.
- En el 10 % de los hogares trabajan ambos.

Se elige una familia al azar, ¿cuál es la probabilidad de que ninguno de los cónyuges trabaje?

El espacio muestral Ω asociado a este experimento está formado por los puntos muestrales ω_1, *trabajan marido y mujer*, ω_2, *sólo trabaja el marido*, ω_3, *sólo trabaja la mujer*, y ω_4, *no trabaja ninguno*. De los resultados de este estudio conocemos las probabilidades de los sucesos $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ y $A \cap B = \{\omega_1\}$, siendo

$$p(A) = 0.70\,, \quad p(B) = 0.20 \quad \text{y} \quad p(A \cap B) = 0.10\,.$$

Con estos datos queremos calcular la probabilidad del suceso $C = \{\omega_4\}$. Teniendo en cuenta que el suceso contrario a *no trabajar ninguno* de los cónyuges es *trabajar al menos uno de ellos*, obtenemos:

$$p(C) = 1 - p(A \cup B) = 1 - [p(A) + p(B) - p(A \cap B)] = 0.20\,.$$

Es decir, en el 20 % de los hogares de dicha población no trabajan ni el marido ni la mujer.

Aunque no se ha explicitado, la σ-álgebra con la que estamos trabajando es $\mathcal{P}(\Omega)$, ya que todos los sucesos elementales tienen asignada una probabilidad, es decir, pertenecen a la σ-álgebra. Así,

$$\begin{array}{lcl} p(\omega_1) & = & p(A \cap B) = 0.10\,, \\ p(\omega_2) & = & p[A - (A \cap B)] = p(A) - p(A \cap B) = 0.60\,, \\ p(\omega_3) & = & p[B - (A \cap B)] = p(B) - p(A \cap B) = 0.10 \quad \text{y} \\ p(\omega_4) & = & p(C) = 0.20\,. \end{array}$$

Adviértase que la notación correcta sería $p(\{w_i\})$, pues la probabilidad se define sobre sucesos, en este caso elementales, y no sobre puntos muestrales. Aun así, por comodidad utilizaremos la notación abreviada. $\triangleleft$

2.5 Algunos tipos de espacios probabilísticos

Para concluir el tema, estudiemos la probabilidad en algunos espacios muestrales de interés.

2.5.1 Probabilidad en un espacio muestral discreto

Con frecuencia se plantean, en el análisis de fenómenos sociales y de otra índole, experimentos aleatorios en los que el conjunto de resultados posibles es finito o infinito numerable. Ejemplos de los primeros podrían ser la posibilidad de que un individuo de características socioeconómicas concretas compre o no un coche (dos resultados posibles), o la observación del número de empleados de la empresa que estarán de baja por enfermedad durante un mes determinado (los resultados posibles son los números naturales menores o iguales que el número de empleados de la empresa); ejemplos de los segundos, la contabilización del número de llamadas telefónicas que se registran anualmente en una gran central o el recuento del número de bacterias en un cultivo. Aunque a estos experimentos son aplicables todos los resultados obtenidos hasta ahora, sus especiales características exigen algunas puntualizaciones.

I. Probabilidad en un espacio muestral finito

Sea el espacio muestral $\Omega = \{\omega_1, \ldots, \omega_n\}$. Supongamos que todos los sucesos elementales son sucesos de la σ-álgebra α, es decir, son sucesos con probabilidad. Como α ha de contener todas las posibles uniones, α ha de ser $\mathcal{P}(\Omega)$. Sea A un suceso cualquiera, es decir, un subconjunto cualquiera de Ω. Para simplificar la notación supondremos que está formado por los primeros k elementos de Ω,

$$A = \{\omega_1, \ldots, \omega_k\} .$$

Entonces, se puede expresar A como unión de sucesos elementales,

$$A = \{\omega_1\} \cup \cdots \cup \{\omega_k\} ,$$

que son disjuntos dos a dos. En consecuencia, y simplificando la notación como será habitual en adelante, obtenemos:

$$p(A) = p(\omega_1) + \cdots + p(\omega_k) .$$

En general, se demuestra que si A es un subconjunto de Ω,

$$p(A) = \sum_{\omega_i \in A} p(\omega_i) .$$

Por tanto, con definir las probabilidades de los sucesos elementales queda perfectamente definida una probabilidad en un espacio muestral finito.

¿Qué propiedades deben cumplir las probabilidades de los n sucesos elementales? Dado que

$$\Omega = \{\omega_1\} \cup \cdots \cup \{\omega_n\}$$

y los sucesos elementales son disjuntos dos a dos,

$$1 = p(\Omega) = p(\omega_1) + \cdots + p(\omega_n) .$$

Por otro lado, dichas probabilidades (como siempre) serán no negativas.

En definitiva, *toda asignación de probabilidad a los sucesos elementales en un espacio muestral finito que cumpla:*

1) $p(\omega_i) \geq 0$ *para* $i = 1, 2, \ldots, n$;

2) $p(\omega_1) + \cdots + p(\omega_n) = 1$

permite construir una probabilidad sobre $\mathcal{P}(\Omega)$ *definida como*

$$p(A) = \sum_{\omega_i \in A} p(\omega_i)$$

para todo subconjunto A *de* Ω.

El lector puede comprobar que la asignación de probabilidad así obtenida verifica la definición de KOLMOGOROV, pudiendo considerarse unos axiomas de KOLMOGOROV transformados para un espacio muestral finito.

Ejemplo 2.4. Al experimento de tirar un dado le podemos asignar una probabilidad con tal de dar a cada suceso elemental un valor de probabilidad positivo y cuya suma sea la unidad; por ejemplo, si $p(1) = p(2) = 1/4$ y el resto de las caras tienen probabilidad $1/8$, tendremos definida la probabilidad de cualquier otro suceso. Así, $p(\text{par}) = p(\{2,4,6\}) = p(2) + p(4) + p(6) = 1/4 + 1/8 + 1/8 = 6/8$; igualmente puede obtenerse la probabilidad de *impar* o de cualquier otro suceso de interés.

Si el dado no está cargado, no parece razonable esta asignación de probabilidad. En este caso sería mas lógico asignar a todas las caras la misma probabilidad:

$$p(1) = p(2) = \cdots = p(6) = K \ .$$

Dado que la suma de las probabilidades de cada cara tiene que ser la unidad, el valor de K será $1/6$. Calculemos la probabilidad de *par* :

$$p(\{2,4,6\}) = p(2) + p(4) + p(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \ .$$

Esta situación especial (igualdad de probabilidades de todos los sucesos elementales) la abordaremos teóricamente a continuación. ◁

II. Probabilidad en un espacio muestral finito y equiprobable

Cuando el espacio muestral es finito y equiprobable, estamos ante un caso particular del anterior donde todos los sucesos elementales tienen la misma probabilidad, esto es,

$$p(\omega_1) = \cdots = p(\omega_n) \ .$$

Como se tiene que cumplir que

$$p(\omega_1) + \cdots + p(\omega_n) = 1 \ ,$$

entonces,

$$p(\omega_i) = \frac{1}{n} \qquad i = 1, \ldots, n \ .$$

Sea A un suceso cualquiera. Vimos anteriormente que

$$p(A) = \sum_{\omega_i \in A} p(\omega_i) \ .$$

En nuestro caso

$$p(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{k}{n} \ ,$$

siendo k el número de elementos de A.

Como n es el número de todos los casos posibles (todos los resultados simples del experimento) y k el número de éstos que son favorables (resultados simples para los cuales se verifica el suceso), obtenemos la conocida regla de LAPLACE:

$$p(A) = \frac{\text{casos favorables}}{\text{casos posibles}} \ .$$

Esta regla, deducida por LAPLACE y que da lugar al planteamiento clásico de la probabilidad, no es aplicable a cualquier suceso, sino sólo a aquellos pertenecientes a espacios muestrales finitos y equiprobables (algunas situaciones en juegos de azar, estudio de proporciones en una población, etc.)[3].

La utilización de esta regla para el cálculo de probabilidades plantea dos problemas. En primer lugar, se debe garantizar la equiprobabilidad de los sucesos elementales; en segundo lugar, hallar el número de casos posibles y el de casos favorables al suceso cuya probabilidad se quiere obtener.

El primer problema es, tal vez, el más delicado, ya que no existe un método general que permita abordarlo. Por ejemplo, si se lanza una moneda cargada dos veces, la regla de LAPLACE no es aplicable. Si la moneda es perfecta, puede aplicarse si escribimos el espacio muestral como

$$\Omega_1 = \{(c,c),(c,z),(z,c),(z,z)\} ,$$

pero no si lo escribimos como

$$\Omega_2 = \{2\text{caras}, 1\text{cara}, 0\text{caras}\} .$$

Parece razonable que los cuatro sucesos elementales de Ω_1 tengan la misma probabilidad, 1/4 (la formalización de la equiprobabilidad de los cuatro sucesos anteriores, que veremos en el siguiente capítulo, no es sencilla).

Para resolver el segundo problema se han desarrollado algunas técnicas algebraicas que reciben el nombre genérico de Análisis Combinatorio. No es nuestra intención detenernos en un estudio pormenorizado de las mismas y únicamente describiremos al final de este capítulo las nociones y resultados más importantes. El lector que desee abordar estos conceptos de manera más amplia deberá recurrir a textos que los desarrollen con mayor detalle[4].

Ejemplo 2.5. Elijamos al azar una carta de una baraja española. Si suponemos que no están marcados, cada uno de estos cuarenta naipes tendrá la misma probabilidad de ser elegido, 1/40. ¿Cuál es la probabilidad de *elegir oros*?, ¿y de *elegir rey*?

Podemos aplicar la regla de LAPLACE, contabilizando el número de casos favorables para cada suceso. Así,

$$p(O) = p(\text{elegir oros}) = 10/40$$

y

$$p(R) = p(\text{elegir rey }) = 4/40 .$$

Igualmente calcularíamos la probabilidad de *elegir un rey que no sea de oros*,

$$p(R \cap \overline{O}) = 3/40 . \triangleleft$$

3. La regla de LAPLACE se puede aplicar en un contexto ligeramente mas amplio, aquel en el que exista una partición de Ω formada por sucesos equiprobables. En este caso, podríamos calcular solamente la probabilidad de ciertos sucesos, los que se puedan poner como unión de los anteriores.

4. Un enfoque algebraico del Análisis Combinatorio puede verse en MARTÍNEZ SALAS, J., *Elementos de Matemáticas*, Gráficas Andrés Martín, Valladolid, 1977; los textos de FELLER, W., *Introducción a la Teoría de Probabilidades y sus aplicaciones*, vol. I, Limusa-Wiley, México, 1975 y de KALBFLEISCH, J. G., *Probabilidad e Inferencia Estadística*, vol. 1, AC, Madrid, 1984, desarrollan estos conceptos desde el punto de vista del cálculo probabilístico.

Ejemplo 2.6. Lanzamos dos dados normales y vemos el resultado. Los puntos del espacio muestral asociado Ω_1 son:

$$\begin{array}{cccc} (1,1) & (1,2) & \dots & (1,6) \\ (2,1) & (2,2) & \dots & (2,6) \\ \vdots & \vdots & \vdots & \vdots \\ (6,1) & (6,2) & \dots & (6,6) \end{array}$$

Obsérvese que, al describir de esta forma los puntos muestrales, estamos considerando los dados distinguibles. En caso contrario, los puntos muestrales que forman Ω_2 serán:

$$\begin{array}{cccc} (1,1) & (1,2) & \dots & (1,6) \\ & (2,2) & \dots & (2,6) \\ & & \ddots & \vdots \\ & & & (6,6) \end{array}$$

No parece insensato asignar a todos los sucesos elementales de Ω_1 una misma probabilidad. Dado que hay 36 posibles resultados (variaciones con repetición de seis elementos tomados de dos en dos), la probabilidad de cada uno de ellos es $1/36$. Esta suposición de equiprobabilidad no tendría sentido si no hubiésemos considerado los dados distinguibles, es decir, si trabajásemos con el espacio muestral Ω_2. Bajo el supuesto de equiprobabilidad, calcularemos la probabilidad de cualquier suceso mediante la regla de LAPLACE. Por ejemplo, ¿cuál será la probabilidad de *no obtener una pareja*? El número de casos favorables es 30 (variaciones sin repetición de seis elementos tomados de dos en dos). Por tanto:

$$p(\text{no obtener pareja}) = \frac{30}{36} .$$

Igualmente calcularíamos la probabilidad de *obtener al menos un seis*, $11/36$, o la probabilidad de que *la suma de puntuaciones sea par*. En este caso, los resultados favorables son todos los formados por dos puntuaciones pares o por dos impares (en ambos casos, variaciones con repetición de tres elementos tomados de dos en dos).

$$p(\text{obtener suma par}) = \frac{18}{36} . \triangleleft$$

III. Probabilidad en un espacio muestral infinito numerable

Ciertamente, el planteamiento clásico de la probabilidad no resuelve todos los problemas que surgen en el estudio de los experimentos aleatorios, pues, evidentemente, no todos los espacios pueden considerarse finitos ni todos los sucesos equiprobables. Por ejemplo, la probabilidad de elegir un número impar en el conjunto de los números naturales o la de que una persona viva 80 años no pueden calcularse mediante la regla de LAPLACE. En el primer caso, el número posible de resultados es infinito; en el segundo, los resultados posibles no pueden considerarse equiprobables.

Sea $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ un espacio muestral con una infinidad numerable de resultados posibles. Supongamos, como en el caso finito, que la σ-álgebra es

$\mathcal{P}(\Omega)$. Cualquier suceso A puede expresarse como unión de los sucesos elementales contenidos en el mismo:

$$A = \bigcup_{\omega_i \in A} \{\omega_i\}$$

con lo que

$$p(A) = \sum_{\omega_i \in A} p(\omega_i) \, .$$

Por tanto, como ya ocurriera en el caso finito, conocida la probabilidad de los sucesos elementales, puede calcularse la de cualquier suceso.

¿Qué condiciones deben cumplir las probabilidades de los sucesos elementales? Puesto que

$$\Omega = \bigcup_{i=1}^{\infty} \{\omega_i\} \, ,$$

se tiene que

$$1 = p(\Omega) = p\left(\bigcup_{i=1}^{\infty} \{\omega_i\}\right) \, .$$

Luego

$$\sum_{i=1}^{\infty} p(\omega_i) = 1 \, .$$

Consecuentemente, *podemos definir una probabilidad en un espacio muestral infinito numerable con* $\alpha = \mathcal{P}(\Omega)$ *como*

$$p(A) = \sum_{\omega_i \in A} p(\omega_i)$$

donde A *es un suceso cualquiera y las probabilidades* $p(\omega_i)$ *verifican*

1. $p(\omega_i) \geq 0$ *para todo* $\omega_i \in \Omega$;
2. $\sum_{i=1}^{\infty} p(\omega_i) = 1$.

Se comprueba que la probabilidad así definida cumple los axiomas de KOLMOGOROV.

Cuando el espacio es infinito numerable no cabe la suposición de equiprobabilidad. No podríamos asignar a todos los sucesos elementales una probabilidad igual y positiva, porque su suma no sería finita.

Ejemplo 2.7. El número de enfermos cardiovasculares asistidos diariamente en el servicio de urgencias del hospital de una gran ciudad se modeliza mediante la siguiente asignación de probabilidad:

$$p(\text{asistir a } k \text{ enfermos}) = \left(\frac{1}{2}\right)^{k+1} \qquad k = 0, 1, 2, \ldots$$

En esta situación, el espacio muestral asociado está formado por todos los números naturales, $\mathbf{N}$, incluido el cero. Dado que conocemos la probabilidad de cada uno de los

sucesos elementales, la σ-álgebra con la que trabajamos es $\mathcal{P}(\mathbb{N})$. Esta asignación de probabilidad es realmente una probabilidad. Por un lado, a cada suceso elemental se le asocia un valor positivo; por otro, la suma de todas estas probabilidades es la unidad[5],

$$\sum_{i=0}^{\infty} p(\text{asistir a } i \text{ enfermos}) = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} = 1 .$$

A partir de aquí, podemos hallar la probabilidad de cualquier suceso. Así, la probabilidad de *asistir al menos a un enfermo* será:

$$p(\text{asistir al menos a un enfermo}) = 1 - p(\text{asistir a 0 enfermos}) = 1 - \left(\frac{1}{2}\right)^{0+1} = \frac{1}{2}$$

o la de *asistir a lo sumo a 10 enfermos*[6],

$$\sum_{i=0}^{10} p(\text{asistir a } i \text{ enfermos}) = \sum_{i=0}^{10} \left(\frac{1}{2}\right)^{i+1} = 1 - \left(\frac{1}{2}\right)^{11} = 0.9986 .$$

De lo anterior se deduce que, si bien parece inadecuado asignar a este experimento un número infinito numerable de resultados (existe una limitación tanto temporal como demográfica), en la práctica esto no será relevante, pues la probabilidad de que acudan un número elevado de enfermos a urgencias es casi nula. ◁

2.5.2 Probabilidad en un espacio muestral continuo

Los tres tipos de espacios muestrales estudiados en los apartados anteriores hacen referencia a un conjunto finito o infinito numerable de resultados posibles que, de alguna manera, están *separados*. Así, tanto los resultados obtenidos al lanzar un dado, $\{1, 2, 3, 4, 5, 6\}$, como al contabilizar el número de enfermos asistidos diariamente en un servicio de urgencias, $\{0, 1, 2, \ldots\}$, son ejemplos de estas situaciones. Vamos a abordar ahora el estudio de un espacio muestral en el que los posibles resultados no están *separados*, sino que forman un conjunto *continuo* de valores.

Así, si elegimos un punto al azar en el intervalo real $(0, 1)$, los posibles resultados de este experimento están *pegados*[7]. Consideremos la σ-álgebra que contiene a

5. Téngase en cuenta que se trata de sumar una progresión geométrica ilimitada, donde tanto el primer término, a_1, como la razón, r, valen $1/2$.

$$S = \frac{a_1}{1 - r} .$$

6. Recuérdese que el valor de la suma de una progresión geométrica limitada, donde a_1 y a_n son su primer y último término y r su razón, es

$$S = \frac{a_1 - a_n r}{1 - r} .$$

7. Si elegimos un punto en $\Omega = (0, 1)$, por ejemplo 0.675, no podemos decir cuál es el resultado anterior o el posterior (responder que dichos resultados son 0.674 y 0.676 es un craso error; entre 0.674 y 0.675 existen ¡infinitos! posibles resultados). Aunque esta propiedad no caracteriza únicamente a los espacios continuos (piense el lector en la elección de números racionales del intervalo $(0, 1)$), describe suficientemente bien lo que en ellos ocurre.

todos los subintervalos (a, b) del espacio muestral $\Omega = (0, 1)$; dicha σ-álgebra contendrá todas las uniones, intersecciones y complementarios de dichos subintervalos. En capítulos posteriores será necesaria una mayor concreción de esta σ-álgebra (σ-álgebra de BOREL), no así en esta ocasión, en la que utilizaremos únicamente una idea intuitiva de la misma.

Definimos la probabilidad de un subintervalo como la longitud del mismo, es decir,

$$p[(a, b)] = \text{long}[(a, b)] = b - a \ .$$

No nos extenderemos en la consideración sobre cómo definir las probabilidades de sucesos que no sean intervalos; éstas se obtienen utilizando los axiomas de KOLMOGOROV y las propiedades que de ellos se deducen.

El lector puede intuir que una aplicación así definida es realmente una probabilidad. Se utiliza frecuentemente cuando tratamos de explicar la elección de un punto en $\Omega = (0, 1)$ y no existe preferencia por ninguna zona del mismo, es decir, la probabilidad se encuentra uniformemente distribuida a lo largo del intervalo.

Se puede realizar una construcción similar cuando el conjunto de resultados posibles es un intervalo no unitario $\Omega = (c, d)$. Entonces, la probabilidad de un subintervalo (a, b) de Ω se definiría como

$$p[(a, b)] = \frac{\text{long}[(a, b)]}{\text{long}[(c, d)]} = \frac{b - a}{d - c}.$$

La probabilidad anterior generaliza, en cierto modo, la regla de LAPLACE para los espacios muestrales continuos. La equiprobabilidad no se entiende aquí referida a los resultados simples (piense el lector que esa probabilidad sería cero y, además, que las probabilidades de los intervalos no se pueden obtener a partir de las de sus puntos muestrales), sino a los subintervalos de igual tamaño.

El ejemplo anterior propone una probabilidad construida a partir de una *medida* de los intervalos, su longitud. De forma similar podríamos definir probabilidades a partir de otros tipos de *medidas*: superficie de un rectángulo, volumen de un prisma, duración de un período de tiempo, etc.

En capítulos posteriores se formalizará la construcción de probabilidades sobre espacios muestrales continuos.

2.6 Ejercicios propuestos

2.1. Demuéstrense las dos leyes de MORGAN:

a) $\overline{A \cup B} = \overline{A} \cap \overline{B}$.
b) $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

2.2. Hállese el número de sucesos contenidos en $\mathcal{P}(\Omega)$ si el espacio muestral Ω está formado por n puntos muestrales.

2.3. Compruébese que la intersección de toda colección finita de sucesos de una σ-álgebra pertenece a la misma.

2.4. Sea A y B dos sucesos de una σ-álgebra α. Demuéstrese que $A - B$ y $A \Delta B$ también pertenecen a α.

2.5. En una urna hay cuatro bolas numeradas del 1 al 4. Se considera el experimento aleatorio consistente en extraer una bola de la urna. Constrúyase la σ-álgebra generada por los sucesos A, *obtener múltiplo de tres*, y B, *obtener número par.*

2.6. En una urna hay dos bolas blancas y tres negras. Las bolas blancas están numeradas del 1 al 2 y las bolas negras del 1 al 3. Si el experimento aleatorio consiste en sacar una bola de la urna y observar su número y color, constrúyase la σ-álgebra generada por los sucesos A, *salir bola blanca* y B, *salir número par.*

2.7. Una σ-álgebra está generada por m átomos. ¿Cuántos sucesos la componen?

2.8. Dado un espacio muestral Ω, pruébese que $\alpha_1 = \{\emptyset, \Omega\}$ y $\alpha_2 = \mathcal{P}(\Omega)$ son σ-álgebras.

2.9. Sea el espacio muestral $\Omega = \mathbb{N}$ y la σ-álgebra $\mathcal{P}(\Omega)$. Definimos la aplicación $p : \mathcal{P}(\Omega) \longrightarrow \mathbb{R}$ tal que $p(A) = 1$ si 0 o 1 pertenecen a A, y $p(A) = 0$ en el resto de los casos. ¿Es p una probabilidad?

2.10. Asígnese a un experimento determinista una probabilidad que explique su comportamiento.

2.11. Razónese si son verdaderas o falsas las siguientes afirmaciones:

a) Si la suma de las probabilidades de dos sucesos es 1, dichos sucesos son complementarios uno de otro.

b) La igualdad $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ únicamente es cierta cuando $A \cap B$ no es un suceso imposible.

c) Los sucesos A, B y $A \cap B$ tienen las siguientes probabilidades: $p(A) = 0.25$, $p(B) = 0.5$ y $p(A \cap B) = 0.25$.

d) Si en un experimento ocurre el resultado $\omega \in A$, ocurre el suceso A y sólo él.

e) Un suceso y su contrario pueden tener probabilidades cuyo producto sea uno.

2.12. Sean A y B dos sucesos mutuamente excluyentes tales que $p(A) = 0.3$ y $p(B) = 0.6$. Calcúlense las probabilidades de los siguientes sucesos: $\overline{A}$, $\overline{B}$, $A \cap B$, $A \cup B$, $\overline{A} \cap \overline{B}$ y $\overline{A} \cup \overline{B}$.

2.13. Sean A, B y C tres sucesos tales que $A \cap B \subset C$. Demuéstrese que

$$p(\overline{C}) \leq p(\overline{A}) + p(\overline{B}) \ .$$

2.14. Sean A y B dos sucesos tales que $p(A) = 0.3$, $p(B) = 0.6$ y $p(A \cap B) = 0.2$. Hállense las probabilidades de los siguientes sucesos: $\overline{A}$, $\overline{B}$, $A \cap \overline{B}$, $\overline{A} \cap B$, $A \cup B$, $\overline{A} \cap \overline{B}$ y $\overline{A} \cup \overline{B}$.

2.15. Sean A, B y C tres sucesos. Demuéstrese que $p(A \cup B) \geq 1 - p(\overline{A}) - p(\overline{B})$.

2.16. Una estadística asegura que el 20 % de los habitantes de una ciudad compra habitualmente discos de música clásica, el 30 % de rock y el 15 % de jazz. Asimismo, el 5 % adquiere discos tanto clásicos como de rock, el 7 % de rock y de jazz, el 6 % clásicos y de jazz, y el 1 % de los tres tipos. Calcúlese:

a) Porcentaje de personas que compran música clásica y no de jazz.
b) Porcentaje de individuos que compran discos de rock o bien de música clásica y de jazz.
c) Porcentaje de personas que compran discos de jazz.

2.17. En una cierta empresa se ha elaborado un informe sobre las actividades que realizan los empleados en su tiempo libre. Dicho análisis arroja, entre otros, los siguientes resultados: el 30 % de los trabajadores practica algún deporte, el 25 % dedica varias horas semanales a la lectura y únicamente el 10 % tiene ambas aficiones.

a) Determínese el porcentaje de trabajadores que sólo practican deporte en sus ratos de ocio.
b) Calcúlese el porcentaje de empleados que ni leen ni realizan actividades deportivas.

2.18. Demuéstrese que las asignaciones de probabilidad a espacios muestrales finitos e infinitos numerables (véase subsección 2.5.1, I y III) verifican los tres axiomas de KOLMOGOROV.

2.19. En una urna hay un total de 12 bolas entre blancas y negras. Sabiendo que la probabilidad de sacar una bola negra es el triple de la de blanca, ¿cuántas bolas blancas y negras hay en la urna?

2.20. Se ha trucado un dado para que la probabilidad de cada cara sea proporcional a la puntuación de la misma. ¿Cuál es la probabilidad de que no salga el seis? ¿Y la de que salga par?

2.21. Sea $\Omega = \{0, 1\}$ y $p : \mathcal{P}(\Omega) \to \mathbb{R}$ tal que $p(i) = 0.3^i \cdot 0.7^{k-i}$ para $i = 0, 1$. Hállese el valor de k para que p sea una probabilidad.

2.22. Una recién casada está decorando el salón de su nueva residencia. Para ello cuenta, entre otros elementos, con tres cuadros que sólo puede colocar en cada una de las tres paredes disponibles al efecto. Una vez tomada la decisión sobre qué pared será la más adecuada para cada cuadro, y sin comunicársela a su marido, requiere su opinión. ¿Cuál es la probabilidad de que el esposo coincida con el criterio de su mujer en la disposición de al menos un cuadro?

2.23. En un concurso se ha producido un cuádruple empate y el jurado ha decidido otorgar un premio distinto, aunque de idéntico valor económico, a cada uno de los ganadores. Si los cuatro tienen preferencias diferentes y el reparto de premios se realiza al azar, ¿cuál es la probabilidad de que todos queden satisfechos?, ¿y de que ninguno se lleve el premio deseado?

2.24. El encargado de mantenimiento de un hospital tiene que colocar placas con los nombres de cinco médicos en las puertas de sus correspondientes despachos. Como ha olvidado a quién pertenece cada uno, decide hacerlo aleatoriamente.

a) ¿Cuál es la probabilidad de que el doctor Fernández tenga bien colocada su placa?

b) Hállese la probabilidad de que exactamente dos médicos tengan identificados sus despachos correctamente.

2.25. Sea el espacio muestral $\Omega = \{1, 3, 5, \ldots\}$, formado por los números impares positivos. Se define la probabilidad de un suceso elemental como:

$$p(i) = k \cdot \left(\frac{1}{3}\right)^i \qquad i = 1, 3, 5, \ldots$$

a) ¿Qué valor tiene que tomar k?

b) Hállese la probabilidad del suceso $A = \{7, 9, 11, \ldots\}$.

2.26. Dado el espacio muestral $\Omega = \{1, 2, \ldots\}$, se considera la aplicación p de $\mathcal{P}(\Omega)$ en $\mathbb{R}$ tal que $p(k) = k \cdot e^{-k}$ donde $k = 1, 2, \ldots$ ¿Es p una probabilidad? (Sugerencia: recuerde el lector que para sumar una serie aritmético-geométrica ilimitada, $S = \sum_{i=1}^{\infty} a_i/b_i$, donde a_i y b_i son los términos generales de una serie aritmética y geométrica respectivamente, se calcula $S - \frac{1}{r} \cdot S$, donde r es la razón de la serie geométrica $\{b_i\}_{i=1}^{\infty}$.)

2.27. Se elige un punto al azar en el intervalo $(0, 1)$ (véase subsección 2.5.2). Calcúlese la probabilidad de elegir un punto concreto, por ejemplo el 0.5. (Sugerencia: el punto 0.5 se puede poner como complementario de la unión de los subintervalos $(0, 0.5)$ y $(0.5, 1)$.)

2.28. Dado el espacio muestral $\Omega = (0, 1)$, se define la probabilidad de un subintervalo (a, b) como el área determinada por dicho subintervalo y la recta $y = 1$. Hállese la probabilidad del subintervalo $(1/4, 1/2)$.

Apéndice: nociones de Combinatoria

Reseñaremos a continuación ciertos conceptos y resultados del Análisis Combinatorio que resultarán de utilidad en el cálculo de probabilidades.

Variaciones

Disponemos de m elementos, $\{a_1, \ldots, a_m\}$, distintos. Llamaremos *variaciones* de estos m elementos tomados de n en n $(n \leq m)$ a los distintos grupos de n elementos formados con los m, considerando dos grupos distintos si difieren en algún elemento o en el orden de los mismos.

En resumen, las variaciones se distinguen por:

- No se repiten los elementos.
- Influye el orden.
- En cada grupo no están incluidos todos los m elementos, salvo que $n = m$.

Se denota por $V_{m,n}$ al número de variaciones de m elementos tomados de n en n, siendo dicho número igual a

$$V_{m,n} = m \cdot (m-1) \cdot (m-2) \cdots (m-n+1) \, .$$

Ejemplo. En una carrera en la que participan diez caballos, ¿de cuántas formas distintas se pueden establecer los cuatro primeros puestos?

En las posibles clasificaciones:

- Cada caballo sólo puede aparecer una vez.
- Influye el orden de llegada.
- Se incluyen sólo cuatro de los diez caballos.

Por tanto se trata de variaciones sin repetición, cuyo número será:

$$V_{10,4} = 10 \cdot 9 \cdot 8 \cdot 7 = 5040 \text{ clasificaciones.} \triangleleft$$

Permutaciones

Dados m elementos distintos, $\{a_1, \ldots, a_m\}$, llamamos *permutaciones* de m elementos a todas las posibles ordenaciones de éstos.

En resumen, las permutaciones se distinguen por:

- No se repiten los elementos.
- Influye el orden.
- En cada grupo están incluidos todos los elementos.

Su número se denota por P_m, coincidiendo, evidentemente, con las variaciones de m elementos tomados de m en m. Es decir:

$$P_m = V_{m,m} = m \cdot (m-1) \cdots (m-m+1) = m! \, .$$

Ejemplo. ¿Cuántas palabras distintas de cuatro letras se pueden formar con la palabra «musa»?

En cada palabra:

- No se repiten las letras.
- Influye el orden de colocación de las mismas.
- Las cuatro letras forman parte de la misma.

Por tanto, se trata de permutaciones, cuyo número será $P_4 = 4! = 24$ palabras. $\triangleleft$

Combinaciones

Dados m elementos distintos, $\{a_1, \ldots, a_m\}$, llamamos *combinaciones* de m elementos tomados de n en n ($n \leq m$) a los distintos grupos de n elementos formados con los m, considerando dos grupos distintos si difieren en algún elemento.

En resumen, las combinaciones se distinguen por:

- No se repiten los elementos.
- No influye el orden.
- En cada grupo no están incluidos todos los m elementos, salvo que $n = m$.

Denotaremos por $C_{m,n}$ al número de combinaciones de m elementos tomados de n en n:

$$C_{m,n} = \binom{m}{n} = \frac{m!}{(m-n)! \cdot n!} .$$

Recordemos una importante propiedad que verifican los números combinatorios:

$$\binom{m}{n} = \binom{m}{m-n} .$$

Ejemplo. Con once empleados, ¿cuántos comités de empresa de cinco personas pueden formarse?

En cada comité:

- No se repiten las personas.
- No influye el orden de las mismas.
- Sólo están incluidas cinco de las once personas.

Por tanto, se trata de combinaciones sin repetición, cuyo número será:

$$C_{11,5} = \binom{11}{5} = \frac{11!}{6! \cdot 5!} = 66 \text{ comités.} \triangleleft$$

Variaciones con repetición

Se tienen m elementos distintos, $\{a_1, \ldots, a_m\}$. Llamaremos *variaciones con repetición* de m elementos tomados de n en n a los distintos grupos de n elementos formados con los m, repetidos o no, considerando dos grupos distintos si difieren en algún elemento, en el número de veces que se repite cada uno de ellos o en el orden de los mismos.

En resumen, distinguiremos las variaciones con repetición por:

- Se pueden repetir los elementos.
- Influye el orden.
- En cada grupo no tienen por qué estar incluidos todos los m elementos.

Designaremos por $VR_{m,n}$ el número de variaciones con repetición de m elementos tomados de n en n:

$$VR_{m,n} = m^n \ .$$

Ejemplo. En el alfabeto Morse de puntos y rayas, ¿cuántos caracteres pueden formarse empleando a lo sumo tres signos?

En cada carácter:

- Los puntos y rayas pueden repetirse.
- Influye el orden en que se disponen.
- No tienen por qué entrar los dos elementos, punto y raya.

En consecuencia, el número de caracteres que podemos formar es:

$$\begin{array}{ll} \text{Con 3 signos,} & VR_{2,3} = 2^3 = 8. \\ \text{Con 2 signos,} & VR_{2,2} = 2^2 = 4. \\ \text{Con 1 signo,} & VR_{2,1} = 2^1 = 2. \end{array}$$

En definitiva, el número total de caracteres es: $8 + 4 + 2 = 14$.◁

Permutaciones con repetición

Dados m elementos, $\{a_1, \ldots, a_m\}$, entre los cuales hay α iguales entre sí y distintos al resto, β iguales entre sí y distintos al resto, ..., y λ iguales entre sí y distintos al resto, se llaman *permutaciones con repetición* de estos m elementos a cada una de las distintas formas de ordenarlos.

En resumen, las permutaciones con repetición se distinguen por:

- En cada grupo están incluidos todos los elementos, repetidos el mismo número de veces.
- Influye el orden.

Se denota $P_m^{\alpha,\beta,\ldots,\lambda}$ al número de permutaciones con repetición:

$$P_m^{\alpha,\beta,\ldots,\lambda} = \frac{m!}{\alpha! \cdot \beta! \ldots \lambda!} \ .$$

Ejemplo. ¿Cuántas palabras distintas es posible formar con todas las letras de la palabra «estadística»?

En cada palabra:

- Se incluyen las mismas letras, repetidas el mismo número de veces.
- Influye el orden de las letras.

Teniendo en cuenta que las letras «s», «t», «a» e «i» se incluyen dos veces, el número de palabras se contabilizan como las permutaciones con repetición de once elementos donde cuatro de ellos se presentan dos veces y el resto una vez, es decir:

$$P_{11}^{2,2,2,2,1,1,1} = \frac{11!}{2! \cdot 2! \cdot 2! \cdot 2! \cdot 1! \cdot 1! \cdot 1!} = 2494800 \text{ palabras.} \quad \triangleleft$$

Combinaciones con repetición

Dados m elementos distintos, $\{a_1, \ldots, a_m\}$, llamamos *combinaciones con repetición* de m elementos tomados de n en n a los distintos grupos de n elementos formados con los m, repetidos o no, considerando dos de ellos distintos si difieren en algún elemento o en el número de veces que se repiten los mismos.

En resumen, las permutaciones con repetición se distinguen por:

- Se pueden repetir los elementos.
- No influye el orden.
- En cada grupo no tienen por qué estar incluidos todos los m elementos.

Denotaremos por $CR_{m,n}$ al número de combinaciones con repetición de m elementos tomados de n en n, siendo dicho número:

$$CR_{m,n} = \binom{m+n-1}{n}.$$

Ejemplo. ¿Cuántas fichas tiene un dominó?
En cada ficha, formada por dos puntuaciones del 0 al 6:

- Se puede repetir la puntuación.
- No influye el orden.
- No están todas las 7 posibles puntuaciones.

Se trata de combinaciones con repetición de siete elementos tomados de dos en dos, es decir:

$$CR_{7,2} = \binom{7+2-1}{2} = \binom{8}{2} = \frac{8!}{6! \cdot 2!} = 28 \text{ fichas.} \quad \triangleleft$$

Ejercicios propuestos

1. Una empresa de reciente creación encarga a un diseñador gráfico la elaboración de su logotipo, indicándole que tiene que seleccionar exactamente tres colores de una lista de seis. ¿Entre cuántos grupos de colores puede decidirse el diseñador?

2. ¿De cuántas formas distintas se puede elegir al delegado y al subdelegado de una clase constituida por cincuenta alumnos?

3. Juan no puede salir a jugar hasta que no ordene su cuarto. Una de las baldas de su estantería está destinada a sus quince cuentos. Si siempre dispone Caperucita Roja en el primer lugar y Los Tres Cerditos en el tercero, ¿cuántas colocaciones distintas puede realizar?

4. Un supermercado organiza una rifa, siendo el premio una botella de cava para todas aquellas papeletas cuyas dos últimas cifras coincidan con las correspondientes al número premiado en el sorteo de Navidad. Suponiendo que todos los boletos tienen cuatro cifras y que existe un único boleto de cada numeración, ¿cuántas botellas repartirá el supermercado?

5. En una tienda de regalos hay relojes de arena con cubetas de colores, no existiendo ninguna diferencia de forma entre las dos que componen cada reloj. Sabiendo que hay cuatro colores posibles y que el color de ambos recipientes puede coincidir, ¿cuántos modelos distintos de reloj de arena se venden en el establecimiento?

6. En una partida de parchís gana aquel jugador que antes consiga *meter en casa* sus cuatro fichas del mismo color. Si son cuatro los jugadores y la partida continúa hasta que todos hayan completado el recorrido, ¿cuántos órdenes distintos hay para la entrada de las dieciséis fichas en casa?

Capítulo 3

PROBABILIDAD CONDICIONADA E INDEPENDENCIA ESTADÍSTICA

3.1 Probabilidad condicionada: definición y propiedades

Supongamos que se tiene un espacio probabilístico asociado a un experimento aleatorio. Disponemos de cierta información y queremos hallar, teniendo en cuenta dicha información, cuál es la probabilidad de un suceso cualquiera. Más concretamente, sabemos que ha ocurrido un determinado suceso y tratamos de hallar la probabilidad de otro, probabilidad que estará influida, o no, por la ocurrencia del primero. Un sencillo ejemplo pondrá de manifiesto la necesidad de definir un nuevo concepto.

Ejemplo 3.1. Realizamos el experimento aleatorio de lanzar un dado normal y observar la puntuación obtenida. La naturaleza del experimento nos permite asumir la hipótesis de equiprobabilidad; es decir, es razonable asignar probabilidades idénticas a cada una de las caras. Por ejemplo, la probabilidad de *obtener un uno* o la de *obtener un dos* será $1/6$.

$$p(\text{obtener un uno}) = \frac{1}{6} \qquad p(\text{obtener un dos}) = \frac{1}{6}\,.$$

De igual forma, calcularíamos la probabilidad de un suceso compuesto, bien aplicando la regla de LAPLACE, bien sumando las probabilidades de todos los sucesos elementales que están contenidos en el mismo. Así, la probabilidad de *obtener un uno o un dos* será $2/6$.

$$p(\text{obtener un uno o un dos}) = \frac{2}{6}\,.$$

Si se ha realizado el experimento y no vemos su resultado, pero un observador nos informa que ha salido *par*, parece lógico que cambiemos la probabilidad de *obtener un uno*. En concreto, la ocurrencia de *par* imposibilita la de *obtener un uno*, debiendo reasignarle una probabilidad nula.

$$p(\text{obtener un uno habiendo obtenido par}) = 0\,.$$

Por otra parte, el conocimiento de que ha salido *par* también nos obligará a recalcular la probabilidad de *obtener un dos*. Ahora, ya no contamos con seis posibles resultados

$(1,2,3,4,5,6)$, sino sólo con tres $(2,4,6)$. Dado que parece lógico seguir manteniendo la hipótesis de equiprobabilidad (de hecho, la información recibida no altera las posibilidades relativas de obtener uno cualquiera de los tres posibles resultados), la probabilidad de *obtener un dos* sería, ahora, $1/3$.

$$p(\text{obtener un dos habiendo obtenido par}) = \frac{1}{3}\,.$$

El mismo razonamiento nos permitirá calcular la probabilidad de *obtener un uno o un dos* habiendo obtenido *par*:

$$p(\text{obtener un uno o un dos habiendo obtenido par}) = \frac{1}{3}\,.$$

Observe el lector que la incorporación de una cierta información, en este caso sabemos que ha ocurrido *par*, influye en la probabilidad de los sucesos. Así, el suceso *obtener un uno* ha visto disminuida su probabilidad (de hecho, se ha anulado), mientras que la de *obtener un dos* ha aumentado. En cambio, la probabilidad del suceso *obtener un uno o un dos* no se ha alterado por la ocurrencia de *par*.

Téngase en cuenta también que, sabiendo que ha ocurrido *par*, la probabilidad de *obtener un dos* y la de *obtener un uno o un dos* coinciden. Ello se debe a que las probabilidades de estos sucesos dependen de un mismo hecho, que es la aparición, o no, del *dos*, es decir, de la intersección de dichos sucesos, $\{2\}$ y $\{1,2\}$, con el suceso que ha ocurrido, $\{2,4,6\}$. En este sentido, también los sucesos $\{2,3\}$, $\{2,5\}$, $\{1,2,3\}$, $\{1,2,5\}$, $\{2,3,5\}$ y $\{1,2,3,5\}$ tendrán la misma probabilidad que los anteriores cuando sabemos que ha ocurrido $\{2,4,6\}$, es decir, $1/3$.

Por último, la probabilidad de *obtener un uno o un dos*, habiendo obtenido *par*, que denotaremos por $p(\{1,2\} \mid \{2,4,6\})$, puede expresarse como:

$$p(\{1,2\} \mid \{2,4,6\}) = \frac{1}{3} = \frac{1/6}{3/6} = \frac{p(\{1,2\} \cap \{2,4,6\})}{p(\{2,4,6\})}\,.$$

Lo anterior corrobora el hecho de que la probabilidad de un suceso, sabiendo que ha ocurrido otro, depende de la intersección de ambos. Además, presenta la ventaja de poder expresar dicha probabilidad en función de las probabilidades del espacio probabilístico inicial. ◁

Generalizando las ideas expuestas en el ejemplo anterior, podemos dar la siguiente definición.

Definición 3.1 *Sea (Ω, α, p) un espacio probabilístico y B un suceso de α de probabilidad no nula, $p(B) > 0$. Sea A un suceso cualquiera de α; denominamos probabilidad del suceso A condicionada por que haya ocurrido el suceso B o, sencillamente, probabilidad de A condicionada por B, al número $p(A \mid B)$ definido por*

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}\,.$$

Con esta definición tratamos de medir la ocurrencia de un suceso A cuando tenemos una cierta información (sabemos que ha ocurrido B). De ahí que exijamos

que la probabilidad de B sea no nula (B es un suceso que puede ocurrir). Por otro lado, no podríamos definir el cociente entre $p(A \cap B)$ y $p(B)$ si $p(B)$ fuese cero[1].

Si entendemos una probabilidad como la medida relativa de la ocurrencia de un suceso, la probabilidad de A condicionada por B sería la medida de A dentro del suceso B, es decir, la medida de la parte de A que está en B, como muestra la definición. De hecho, la probabilidad de A condicionada por B depende de la probabilidad de la intersección de A y B, dividiéndose por la probabilidad de B a los únicos efectos de normalización, de forma que B sea ahora un suceso seguro. En este sentido, el lector puede comprobar que dos sucesos, cuyas intersecciones con el suceso condicionante coinciden, tienen la misma probabilidad condicionada respecto de éste.

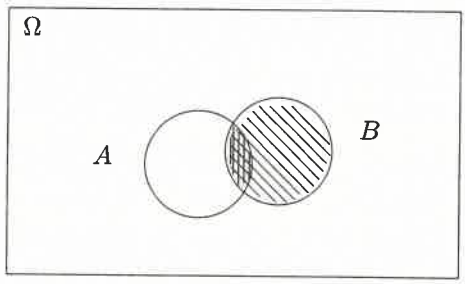


FIG. 3.1.

Demostraremos, a continuación, que lo que hemos designado como *probabilidad condicionada* es realmente una probabilidad.

Proposición 3.1 *Sea (Ω, α, p) un espacio probabilístico y B un suceso de probabilidad no nula. La aplicación $p(\cdot \mid B)$, que a cada suceso A de α le asigna el número $p(A \mid B)$, es una probabilidad sobre el espacio probabilizable (Ω, α).*

Demostración. Comprobemos que una probabilidad condicionada verifica los tres axiomas de probabilidad enunciados en el capítulo anterior. Así,

1. $p(A \mid B) \geq 0$ para todo $A \in \alpha$, pues $p(A \cap B) \geq 0$ y $p(B) > 0$.
2. $p(\Omega \mid B) = p(B \cap \Omega)/p(B) = p(B)/p(B) = 1$.
3. Para toda colección infinita numerable de sucesos de α disjuntos dos a dos, $\{A_i\}_{i=1}^{\infty}$,

$$p\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} p(A_i \mid B) .$$

1. Más adelante veremos cómo definir probabilidades condicionadas por sucesos que, aunque son posibles, tienen probabilidad nula.

En efecto,

$$p\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \frac{p\left[B \cap \left(\bigcup_1^\infty A_i\right)\right]}{p(B)} = \frac{p\left[\bigcup_1^\infty (B \cap A_i)\right]}{p(B)} =$$

$$= \frac{\sum_1^\infty p(B \cap A_i)}{p(B)} = \sum_{i=1}^{\infty} \frac{p(B \cap A_i)}{p(B)} =$$

$$= \sum_{i=1}^{\infty} p(A_i \mid B)\,,$$

donde la tercera igualdad se deduce de que los sucesos de la colección $\{B \cap A_i\}_{i=1}^{\infty}$ son disjuntos dos a dos. ♣

De lo anterior concluimos que la terna $(\Omega, \alpha, p(\cdot \mid B))$ es un espacio probabilístico. En consecuencia, todas las propiedades ya demostradas en el capítulo anterior para una probabilidad, son válidas también para una probabilidad condicionada. Enunciémoslas a continuación:

Corolario 3.1.1 *Sea B un suceso de probabilidad no nula del espacio probabilístico (Ω, α, p); entonces,*

1. $p(\emptyset \mid B) = 0$.
2. $p\left(\bigcup_{i=1}^n A_i \,\middle|\, B\right) = \sum_{i=1}^n p(A_i \mid B)$ *para toda colección finita,* $\{A_i\}_{i=1}^n$, *de sucesos de α disjuntos dos a dos.*
3. $p(\overline{A} \mid B) = 1 - p(A \mid B)$ *para todo* $A \in \alpha$.
4. *Para cualesquiera* $A, C \in \alpha$ *tal que* $A \subset C$, $p(A \mid B) \leq p(C \mid B)$.
5. $0 \leq p(A \mid B) \leq 1$.
6. *Para cualesquiera* $A, C \in \alpha$ *tal que* $A \subset C$, $p[(C - A) \mid B] = p(C \mid B) - p(A \mid B)$.
7. $p[(A \cup C) \mid B] = p(A \mid B) + p(C \mid B) - p[(A \cap C) \mid B]$ *para cualesquiera* $A, C \in \alpha$.
8. $p[(A \cup C) \mid B] \leq p(A \mid B) + p(C \mid B)$ *para cualesquiera* $A, C \in \alpha$.
9. *Para cualesquiera* $A, C, D \in \alpha$,

$$\begin{aligned} p[(A \cup C \cup D) \mid B] &= p(A \mid B) + p(C \mid B) + p(D \mid B) - p[(A \cap C) \mid B] - \\ &\quad p[(A \cap D) \mid B] - p[(C \cap D) \mid B] + p[(A \cap C \cap D) \mid B] \end{aligned}$$

Esta propiedad se generaliza para n sucesos:

$$p\left(\bigcup_{i=1}^{n} A_i \,\middle|\, B\right) = \sum_{i=1}^{n} p(A_i \mid B) - \sum_{\substack{i,j=1 \\ i \neq j}}^{n} p[(A_i A_j) \mid B] + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^{n} p[(A_i A_j A_k) \mid B] + \cdots$$

Aunque este corolario es consecuencia inmediata de la proposición anterior, el lector puede demostrar estos resultados directamente, sin recurrir al hecho de que la probabilidad condicionada es una probabilidad. Por ejemplo, para el primero de ellos resulta:

$$p(\emptyset \mid B) = \frac{p(\emptyset \cap B)}{p(B)} = \frac{p(\emptyset)}{p(B)} = 0 \ .$$

De manera similar comprobaríamos el resto de los resultados.

OBSERVACIONES

1) El concepto de probabilidad condicionada tiene sus antecedentes en el de frecuencia relativa condicionada. Si hemos realizado N experimentos, y los sucesos B y $A \cap B$ han ocurrido n_B y $n_{A\cap B}$ veces, respectivamente, podemos plantear la siguiente igualdad:

$$\frac{n_{A\cap B}}{N} = \frac{n_{A\cap B}}{n_B} \cdot \frac{n_B}{N} \ ,$$

donde $n_{A\cap B}/n_B$ es la proporción de ocurrencias de A cuando ocurrió B, es decir, la frecuencia relativa de A condicionada por B. De aquí, designando este valor por $fr(A \mid B)$, resulta:

$$fr(A \mid B) = \frac{fr(A \cap B)}{fr(B)} \ .$$

2) Tenga en cuenta el lector que el suceso seguro, Ω, no es el único cuya probabilidad condicionada por B es igual a la unidad; lo mismo ocurre con cualquier suceso que contenga a B. Si $A \supset B$,

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B)}{p(B)} = 1 \ .$$

Igualmente, no es el suceso imposible, $\emptyset$, el único con probabilidad nula; también tiene probabilidad nula cualquier suceso incompatible con B, es decir, cualquier suceso contenido en $\overline{B}$. Si $A \subset \overline{B}$,

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p(\emptyset)}{p(B)} = 0 \ .$$

Este hecho puede parecer sorprendente pero, como aclararemos posteriormente en un contexto más amplio, resulta razonable, e incluso necesario a veces, admitir que sucesos distintos del suceso seguro tengan una probabilidad igual a la unidad, y que sucesos distintos del imposible tengan probabilidad nula.

Sin embargo, el problema que aquí se plantea es más amplio y consiste en que sucesos diferentes de Ω (cualquier suceso que contenga a B) son también sucesos seguros, en el sentido que se verifican siempre, y cualquier suceso incompatible con B, es decir, contenido en $\overline{B}$, es un suceso imposible. La explicación a esta aparente paradoja se encuentra en el hecho de que, al condicionar la probabilidad por B, ocurre como si dejasen de ser resultados posibles del experimento aleatorio todos aquellos elementos de Ω que no pertenezcan a B.

3) La observación anterior sugiere una definición alternativa. Podemos definir la probabilidad condicionada sobre el espacio probabilizable (B, α^*), utilizando como espacio muestral al propio B y como σ-álgebra de partes de B la formada por las intersecciones de los sucesos de α con el suceso B, es decir,

$$\alpha^* = \{A^* \in \mathcal{P}(B) \mid A^* = A \cap B \quad \text{donde } A \in \alpha\} .$$

Entonces, la probabilidad condicionada por B se definiría de la siguiente forma: *si A^* es un suceso de α^*, definimos*

$$p(A^* \mid B) = \frac{p(A^*)}{p(B)} .$$

Esta definición es coherente con la definición 3.1 ya que, como A^* está contenido en B, su intersección es A^*.

Esta solución es aparentemente satisfactoria y resulta aceptable cuando se estudian probabilidades condicionadas por un único suceso. Sin embargo, la situación general, en la que se estudian simultáneamente probabilidades condicionadas por diferentes sucesos de la σ-álgebra, nos obligaría a trabajar, a la vez, con diferentes espacios muestrales y σ-álgebras. Por ello, normalmente es preferible considerar que las probabilidades condicionadas se definen sobre el espacio probabilizable (Ω, α), mientras que la identificación de sucesos con subconjuntos del espacio muestral y la determinación de los resultados posibles del experimento se efectúan anteriormente a su realización y, por tanto, sin información previa sobre los resultados del mismo.

4) A partir de un espacio probabilístico (Ω, α, p), podremos definir tantas probabilidades condicionadas como sucesos de probabilidad no nula tenga la σ-álgebra α. En ciertas situaciones, será conveniente considerar conjuntamente toda la familia de probabilidades condicionadas, incluso cuando el suceso condicionante es Ω; en este caso:

$$p(A \mid \Omega) = \frac{p(A \cap \Omega)}{p(\Omega)} = p(A) .$$

De este modo, la probabilidad condicionada podría entenderse como una generalización de la probabilidad.

Veamos, a continuación, una importante consecuencia de la definición de probabilidad condicionada, conocida como **Regla de la multiplicación**.

Proposición 3.2 *Sean n sucesos, $A_1, \ldots, A_n$, tales que $p(A_1 \cap \ldots \cap A_{n-1}) \neq 0$; entonces,*

$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1) \cdot p(A_2 \mid A_1) \cdot \ldots \cdot p(A_n \mid A_1 \cap \ldots \cap A_{n-1}) .$$

Demostración. Realizaremos *inducción matemática* sobre n; es decir, lo demostraremos primero para $n = 2$ y, supuesto que se verifique para un valor particular de n, $n = k$, lo probaremos para $k + 1$.

Para $n = 2$, de la definición de $p(A_2 \mid A_1)$, siendo $p(A_1) \neq 0$, obtenemos:

$$p(A_1 \cap A_2) = p(A_1) \cdot p(A_2 \mid A_1)$$

Supongamos cierto el resultado para un $n = k$ fijo ($k \geq 2$):

$$p(A_1 \cap \ldots \cap A_k) = p(A_1) \cdot p(A_2 \mid A_1) \cdot \ldots \cdot p(A_k \mid A_1 \cap \ldots \cap A_{k-1})$$

y demostrémoslo para $k+1$ cuando $p(A_1 \cap A_2 \cap \ldots \cap A_k) \neq 0$. Evidentemente,

$$p(A_1 \cap \ldots \cap A_k \cap A_{k+1}) = p(A_1 \cap \ldots \cap A_k) \cdot p(A_{k+1} \mid A_1 \cap \ldots \cap A_k) \, .$$

Como suponemos que la regla de la multiplicación se cumple para k, tendremos:

$$p(A_1 \cap \ldots \cap A_k \cap A_{k+1}) =$$
$$= p(A_1) \cdot p(A_2 \mid A_1) \cdot \ldots \cdot p(A_k \mid A_1 \cap \ldots \cap A_{k-1}) \cdot p(A_{k+1} \mid A_1 \cap \ldots \cap A_k)$$

Por tanto, queda demostrado para cualquier $k+1$, partiendo de que el resultado es cierto para k. Puesto que para $k = 2$ se ha comprobado directamente, deduciríamos que para $k = 3$ también es cierto, y así sucesivamente. Como cualquier entero puede alcanzarse con este proceso, la igualdad es cierta para todo valor de n. ♣

3.2 Teorema de la probabilidad total y teorema de Bayes

Los dos resultados que demostraremos en esta sección son herramientas importantes para el cálculo de probabilidades. El primero de ellos recibe el nombre de **Teorema de la probabilidad total.**

Teorema 3.1 *Sea (Ω, α, p) un espacio probabilístico y sea $\{A_i\}_{i=1}^{\infty}$ una partición de Ω formada por una colección infinita numerable de sucesos de probabilidad no nula. La probabilidad de un suceso cualquiera B puede expresarse como*

$$p(B) = \sum_{i=1}^{\infty} p(A_i) \cdot p(B \mid A_i) \, .$$

Demostración. Escribamos el suceso B de la siguiente manera:

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) = \bigcup_{i=1}^{\infty} (B \cap A_i) \, .$$

Teniendo en cuenta que los sucesos $B \cap A_i$ son disjuntos dos a dos, al serlo los A_i, obtenemos:

$$p(B) = \sum_{i=1}^{\infty} p(B \cap A_i) = \sum_{i=1}^{\infty} p(A_i) \cdot p(B \mid A_i) \, . \, ♣$$

Es obvio que el teorema anterior se verifica, también, para una colección finita de sucesos.

Representemos mediante un diagrama de VENN la partición del espacio Ω y el suceso B.

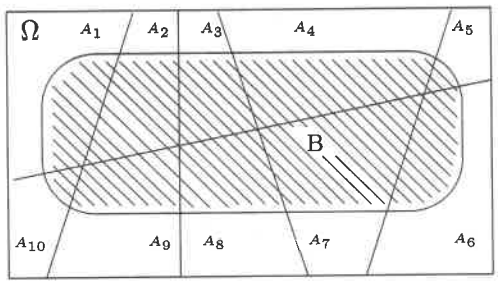


FIG. 3.2.

El teorema nos proporciona la probabilidad de B en función de las probabilidades de los sucesos de la partición y de las probabilidades de B condicionadas por dichos sucesos.

Ejemplo 3.2. Un estudio sociológico clasifica a los individuos de un país según su clase social, estimándose que el 30 % pertenece a la clase alta y el 70 % a la clase baja. Se sabe que el 80 % de los individuos de la clase alta votan al Partido Conservador, mientras que ese porcentaje es sólo del 20 % en la clase baja. ¿Cuál es la probabilidad de que un individuo elegido al azar vote al Partido Conservador?

El espacio muestral asociado al experimento de elegir un individuo y observar la clase a la que pertenece, y si vota o no al Partido Conservador, consta de cuatro resultados posibles: ω_1, *pertenecer a la clase alta y votar conservador*, ω_2, *pertenecer a la clase alta y no votar conservador*, ω_3, *pertenecer a la clase baja y votar conservador*, y ω_4, *pertenecer a la clase baja y no votar conservador*. Denotaremos por A, B y C a los sucesos *pertenecer a la clase alta*, *pertenecer a la clase baja* y *votar conservador*, respectivamente, sucesos que se pueden explicitar por los puntos muestrales que los forman:

$$\begin{array}{rcl} A & = & \{\omega_1, \omega_2\}\,, \\ B & = & \{\omega_3, \omega_4\} \text{ y} \\ C & = & \{\omega_1, \omega_3\}\,. \end{array}$$

El enunciado nos proporciona la siguiente información acerca de la probabilidad de estos sucesos:

$$p(A) = 0.3, \qquad p(B) = 0.7, \qquad p(C \mid A) = 0.8 \quad \text{y} \quad p(C \mid B) = 0.2\,.$$

En virtud del teorema de la probabilidad total, la probabilidad de *votar al Partido Conservador* puede expresarse como:

$$p(C) = p(A) \cdot p(C \mid A) + p(B) \cdot p(C \mid B)$$

que, sustituyendo por los valores correspondientes, resulta:

$$p(C) = 0.38\,.$$

Es decir, el 38 % de los individuos del país vota al Partido Conservador.

El lector puede adivinar la posibilidad de modelizar el ejemplo planteado mediante dos experimentos. El primer experimento sería observar la clase social del individuo; el

otro, observar si vota o no al Partido Conservador. La asignación de probabilidades que exigiría esta modelización es compleja y será abordada en un apartado posterior de este capítulo. ◁

El segundo resultado, de gran importancia, recibe el nombre del reverendo inglés que lo enunció y demostró por primera vez: **Teorema de Bayes**.

Teorema 3.2 *Sea (Ω, α, p) un espacio probabilístico y $\{A_i\}_{i=1}^{\infty}$ una partición de Ω formada por una colección infinita numerable de sucesos de probabilidad no nula. Sea un suceso B tal que $p(B) \neq 0$; entonces,*

$$p(A_j \mid B) = \frac{p(A_j) \cdot p(B \mid A_j)}{\sum_{i=1}^{\infty} p(A_i) \cdot p(B \mid A_i)} \qquad j = 1, \ldots, n, \ldots$$

Demostración. Teniendo en cuenta la definición de probabilidad condicionada,

$$p(A_j \mid B) = \frac{p(A_j \cap B)}{p(B)} = \frac{p(A_j) \cdot p(B \mid A_j)}{\sum_{i=1}^{\infty} p(A_i) \cdot p(B \mid A_i)} ,$$

donde la última igualdad se obtiene al aplicar la regla de la multiplicación en el numerador y el teorema de la probabilidad total en el denominador. ♣

Este resultado, obviamente, se verifica si la colección de sucesos es finita.

El teorema nos proporciona la probabilidad de A_j condicionada por B en función de las probabilidades de los sucesos de la partición y de las probabilidades de B condicionadas por dichos sucesos.

Ejemplo 3.3. Utilizando los datos del ejemplo anterior, podemos estar interesados en calcular la probabilidad de que un individuo, que ha votado al Partido Conservador, pertenezca a la clase alta. En virtud del teorema de Bayes,

$$p(A \mid C) = \frac{p(A) \cdot p(C \mid A)}{p(A) \cdot p(C \mid A) + p(B) \cdot p(C \mid B)} .$$

Sustituyendo los datos iniciales, obtenemos:

$$p(A \mid C) = 0.63 .$$

Es decir, el 63 % de los votantes conservadores pertenece a la clase alta. ◁

Hay ciertas situaciones en las que un suceso B se puede entender como el **efecto** de distintas **causas**, $A_i (i = 1, \ldots, n, \ldots)$. Conocida la probabilidad de cada una de las causas, $p(A_i)$, y la probabilidad de que la ocurrencia de B sea efecto de la causa A_i, $p(B \mid A_i)$, el teorema de Bayes nos proporciona la probabilidad de que, detectado el efecto B, sea A_j la causa que lo provocó, $p(A_j \mid B)$.

El teorema de Bayes tiene una gran importancia en la Estadística; a partir de él se ha desarrollado un enfoque novedoso de la misma: la Estadística bayesiana.

En este enfoque, la probabilidad de un suceso se va afinando con sucesivas incorporaciones de información. Se parte de una probabilidad (probabilidad *a priori*); la asunción de una cierta información nos lleva a reasignar una nueva probabilidad (probabilidad *a posteriori*). En este sentido, asignamos a cada A_i un grado de creencia antes de que se obtenga una cierta evidencia experimental; este grado de creencia es su probabilidad, $p(A_i)$. Ante una determinada evidencia experimental, B, corregimos este grado de creencia, obteniendo una nueva probabilidad para cada A_i, $p(A_i \mid B)$.

Experiencias de Inteligencia Artificial, como el diagnóstico automático en medicina, tienen su base teórica en este teorema. Dicho diagnóstico automático no hace más que repetir el razonamiento del médico a la hora de detectar una enfermedad: analizando los efectos (síntomas), busca la causa que los provocó (enfermedad).

3.3 Independencia de sucesos

La palabra *independencia* se utiliza con frecuencia en el lenguaje cotidiano en relación con los fenómenos aleatorios. Si decimos que el hecho de que suba el precio de la gasolina es independiente del nivel salarial de los funcionarios, estamos expresando nuestro convencimiento de que el mecanismo que rige la formación de los precios de la gasolina y el que determina el sueldo de los funcionarios no están relacionados. Esta idea se traduce en la afirmación probabilística de que el conocimiento de una subida de la gasolina no modificaría las expectativas de mejoras salariales de los funcionarios; éstas serían las mismas, tanto si sabemos que la gasolina ha subido, como si no lo sabemos.

En este mismo sentido decimos que los gastos de una familia dependen de sus ingresos o que la productividad de una empresa es independiente del grado de afiliación sindical de sus empleados.

Vamos a tratar de formalizar esta idea de independencia. Dicho concepto se va a presentar en distintos contextos del Cálculo de Probabilidades y de la Estadística, siempre bajo un sustrato común: la independencia entre sucesos de un mismo espacio probabilístico. Un ejemplo situará el problema en los términos deseados.

Dos jugadores apuestan sobre cuál será la carta elegida al azar en una baraja española normal. Uno de ellos ha apostado por *rey*. Extraída la carta, su rival le comunica que ha salido *espadas* y le ofrece la posibilidad de aumentar o disminuir la cantidad apostada. Teniendo en cuenta la información, ¿es razonable que el jugador modifique su apuesta? Parece evidente (al menos para los jugadores experimentados) que el hecho de saber que ha salido *espadas* no varía las expectativas de que la carta sea un *rey*. Antes de extraer la carta, las posibilidades de acertar que tenía el jugador eran cuatro de cuarenta. A la vista de la información, estas posibilidades siguen siendo las mismas, una de diez. Dicho de otra forma, la ocurrencia de *espadas* no influye, ni positiva ni negativamente, en la de *rey*. Por tanto, no hay ningún motivo razonable para que el jugador cambie su apuesta.

Concretemos el concepto de independencia estadística de sucesos con la siguiente definición.

Definición 3.2 *Dado un espacio probabilístico* (Ω, α, p) *y dados dos sucesos* A *y* B *de* α*, diremos que* A *y* B *son estadísticamente independientes si*

$$p(A \cap B) = p(A) \cdot p(B) \ .$$

Diremos que dos sucesos son estadísticamente dependientes si no son independientes.

Siempre que no se preste a confusión, hablaremos sencillamente de sucesos dependientes o independientes.

Esta definición parece un tanto incongruente con las intuiciones expuestas en el ejemplo anterior. En él, la idea de independencia entre *rey* y *espadas* estaba ligada a que la ocurrencia del suceso *espadas* no influía en la del suceso *rey*; por tanto, esa idea venía expresada en términos de probabilidades condicionadas.

Para sucesos con probabilidad distinta de cero, puede darse una definición alternativa de independencia de sucesos a través de las probabilidades condicionadas que, tal vez, sea más sugerente desde el punto de vista intuitivo y, desde luego, conecta mejor con las ideas expuestas en el citado ejemplo.

Proposición 3.3 *Sea un espacio probabilístico* (Ω, α, p) *y dos sucesos* A *y* B *de* α*, con* $p(A) \neq 0$*; entonces* A *y* B *son independientes si y sólo si*

$$p(B \mid A) = p(B) \ .$$

Demostración. En efecto, si A y B son independientes, aplicando la definición 3.2, obtenemos:

$$p(B \mid A) = \frac{p(A \cap B)}{p(A)} = \frac{p(A) \cdot p(B)}{p(A)} = p(B) \ .$$

Recíprocamente, si se verifica que $p(B \mid A) = p(B)$, aplicando la regla de la multiplicación:

$$p(A \cap B) = p(A) \cdot p(B \mid A) = p(A) \cdot p(B) \ . \clubsuit$$

Cuando uno de los dos sucesos tiene probabilidad nula, la definición de independencia tiene un contenido trivial, como veremos a continuación:

Proposición 3.4 *Si* A *es un suceso de probabilidad nula, entonces* A *es independiente de cualquier suceso* B*. En particular, el suceso imposible,* $\emptyset$*, es independiente de cualquier otro.*

Demostración. Por una parte se tiene que

$$p(A) \cdot p(B) = 0 \cdot p(B) = 0 \ .$$

Por otra, dado que $A \cap B$ está incluido en A, $p(A \cap B) \leq p(A) = 0$, con lo cual

$$p(A \cap B) = 0 \ ,$$

ya que la probabilidad de un suceso nunca puede ser negativa. Así, deducimos que

$$p(A \cap B) = p(A) \cdot p(B) \ .$$

Luego, un suceso A, con $p(A) = 0$, es independiente de cualquier otro suceso B de la σ-álgebra α. ♣

OBSERVACIONES

1) Nótese que las dos proposiciones anteriores constituyen una definición equivalente de independencia; dicho de otra forma, puede afirmarse que, *dado un espacio probabilístico* (Ω, α, p) *y dados dos sucesos* A *y* B *de* α, A *y* B *son independientes si, o bien uno de ellos es de probabilidad nula, o bien* $p(B \mid A) = p(B)$. Esta definición presenta la ventaja de ser más transparente, poniendo en relieve la idea intuitiva de independencia, idea que, por otra parte, está detrás de ambas definiciones. Un sencillo ejemplo ilustrará esta afirmación.

Ejemplo 3.4. Estudios epidemiológicos aseguran que el 10 % de una población ha contraído cierta enfermedad, no habiéndose detectado indicios que permitan afirmar que incida más en uno u otro sexo. Sabiendo que las mujeres forman el 60 % de la población, ¿qué porcentaje representan las que están contagiadas?

Si queremos dar el porcentaje sobre el total de la población, teniendo en cuenta que la enfermedad afecta a las mujeres en la misma medida que a los varones, el 10 % del 60 % de mujeres estará contagiado, es decir, el 6 %. En términos probabilísticos, designando por M y E los sucesos *ser mujer* y *haber contraído la enfermedad*, respectivamente, podemos escribir:

$$p(M \cap E) = p(M) \cdot p(E) = 0.6 \cdot 0.1 = 0.06 \ .$$

Si queremos dar el resultado sobre el total de mujeres, el porcentaje de contagiados entre las mujeres será el mismo que en toda la población, es decir, el 10 %. En lenguaje probabilístico,

$$p(E \mid M) = p(E) = 0.1 \ . \triangleleft$$

2) Como ocurre con todos los conceptos básicos del Cálculo de Probabilidades, el de independencia también tiene sus antecedentes en el campo de las frecuencias. Así, si hemos realizado N veces un experimento aleatorio, podemos decir que dos sucesos A y B son independientes si

$$\frac{n_{A\cap B}}{N} = \frac{n_A}{N} \cdot \frac{n_B}{N} \ ,$$

es decir,

$$fr(A \cap B) = fr(A) \cdot fr(B) \ ,$$

donde $n_{A\cap B}, n_A$ y n_B son el número de veces que se han presentado los sucesos $A \cap B, A$ y B, respectivamente.

Igualmente se puede hablar de independencia de sucesos en términos de frecuencias relativas condicionadas.

Fijemos los anteriores conceptos con el siguiente ejemplo.

Ejemplo 3.5. Realicemos el experimento consistente en elegir una carta al azar de una baraja española no trucada y consideremos los sucesos F, *sacar figura*, R, *sacar rey*, E, *sacar espadas*, y O, *sacar oros*. Nos interesa estudiar la independencia entre los sucesos R y F, R y E, y E y O. En este caso, podemos suponer que el espacio muestral es finito y equiprobable. Utilizando la regla de LAPLACE,

$$p(F) = 12/40, \qquad p(R) = 4/40, \qquad p(E) = 10/40 \quad \text{y} \quad p(O) = 10/40 \, .$$

Si calculamos las probabilidades condicionadas y las comparamos con las anteriores,

$$\begin{aligned} p(R \mid F) = \frac{4}{12} &\neq p(R) = \frac{4}{40} \, , \\ p(R \mid E) = \frac{1}{10} &= p(R) = \frac{4}{40} \quad \text{y} \\ p(E \mid O) = 0 &\neq p(E) = \frac{10}{40} \, . \end{aligned}$$

De aquí se deduce que los sucesos R y E son independientes, mientras que los sucesos R y F son dependientes, al igual que E y O. En concreto, la ocurrencia de *rey* resulta favorecida por la de *figura*, en cambio, la ocurrencia de *espada* resulta perjudicada por la de *oros* (aún más, la ocurrencia de *oros* hace imposible la del suceso *espadas*). Pueden comprobarse los resultados obtenidos, aplicando directamente la primera definición de independencia.

Medite el lector si, intuitivamente, hubiera dado la misma respuesta. ◁

Veamos, a continuación, algunos resultados de interés a propósito de la independencia de sucesos.

Proposición 3.5 *Sea un espacio probabilístico (Ω, α, p) y sean dos sucesos A y B de α. Si A y B son independientes, también lo son A y $\overline{B}$ (en consecuencia, lo serán asimismo $\overline{A}$ y $\overline{B}$).*

Demostración. Expresemos el suceso A como unión de sucesos disjuntos,

$$A = (A \cap B) \cup (A \cap \overline{B}) \, .$$

Entonces,

$$p(A \cap \overline{B}) = p(A) - p(A \cap B) \, .$$

Como A y B son independientes,

$$p(A \cap \overline{B}) = p(A) - p(A) \cdot p(B) \, ,$$

es decir,

$$p(A \cap \overline{B}) = p(A) \cdot [1 - p(B)] = p(A) \cdot p(\overline{B}) \, .$$

Si aplicásemos el resultado obtenido a los sucesos A y $\overline{B}$, demostraríamos que $\overline{A}$ y $\overline{B}$ son independientes. ♣

Corolario 3.5.1 *Si A es un suceso tal que $p(A) = 1$, entonces A es independiente de cualquier suceso B. En particular, el suceso seguro, Ω, es independiente de cualquier otro.*

Demostración. El suceso $\overline{A}$ será independiente de cualquier otro por ser de probabilidad nula. Entonces el suceso A también lo será, por ser el complementario de $\overline{A}$. ♣

Proposición 3.6 *Un suceso A es independiente de sí mismo si y sólo si su probabilidad es cero o uno.*

Demostración. La suficiencia de la condición está demostrada en la proposición 3.4 y en el corolario 3.5.1.

Por otro lado, si A es independiente de A,

$$p(A) = p(A \cap A) = p(A) \cdot p(A) = p(A)^2 .$$

La ecuación anterior sólo se verifica si $p(A)$ es 0 o 1. ♣

Como hemos visto, dos sucesos independientes son aquellos que no se influyen en su ocurrencia, lo cual no indica que no tengan resultados en común (es decir, que no puedan ocurrir simultáneamente). Una ligereza en el lenguaje no nos debe hacer confundir los conceptos de sucesos incompatibles y sucesos independientes. No sólo son conceptos distintos sino que, como se prueba en la proposición siguiente, sólo sucesos muy particulares son a la vez incompatibles e independientes.

Proposición 3.7 *Sean A y B dos sucesos incompatibles e independientes. Entonces, al menos uno de ellos es de probabilidad nula.*

Demostración. En efecto, como A y B son incompatibles,

$$p(A \cap B) = p(\emptyset) = 0 .$$

Además, si A y B son independientes,

$$p(A \cap B) = p(A) \cdot p(B) .$$

En definitiva,

$$p(A) \cdot p(B) = 0$$

y, por tanto, $p(A)$ o $p(B)$ se anulan. ♣

Ejemplo 3.6. Lanzamos un dado perfecto y consideramos los siguientes sucesos:

$$A = \{1, 2\}, \qquad B_1 = \{2, 4, 6\} \quad \text{y} \quad B_2 = \{3, 4\} .$$

Evidentemente,

$$p(A) \cdot p(B_1) = \frac{2}{6} \cdot \frac{3}{6} = \frac{1}{6}$$

y

$$p(A \cap B_1) = p(2) = \frac{1}{6} .$$

Por tanto, A y B_1 son independientes pero no disjuntos, mientras que A y B_2 son disjuntos pero no son independientes, como se puede comprobar. ◁

Por último, es conveniente recalcar que la independencia de sucesos no es una propiedad intrínseca de los mismos, es decir, no está ligada a su naturaleza, sino,

exclusivamente, a las probabilidades que dichos sucesos tienen asignadas. Aclaremos esta idea con el ejemplo siguiente.

Ejemplo 3.7. Consideremos los sucesos R y E, *sacar rey* y *sacar espadas* en una baraja española, respectivamente. Bajo la suposición de que la baraja es normal, ya hemos visto que ambos sucesos son independientes (ejemplo 3.5). Supongamos, ahora, que uno de los jugadores está interesado en la carta *as de copas* e incluye en la baraja, sin que lo aprecien sus contrincantes, otra carta igual. Notemos que dicha carta no pertenece ni al suceso R ni al suceso E. Ahora, el espacio muestral ha variado, componiéndose de 41 cartas, de las cuales dos son iguales. Entonces, la probabilidad de obtener un *as de copas* es el doble que la de obtener otra carta cualquiera. Esto da lugar a las siguientes probabilidades:

$$p(R) = \frac{4}{41}, \qquad p(E) = \frac{10}{41} \quad \text{y} \quad p(R \cap E) = \frac{1}{41},$$

es decir, R y E son, ahora, dependientes. Este ejemplo corrobora la idea de que la independencia no radica en la naturaleza de los sucesos, sino en sus probabilidades. Piense, por tanto, el lector que la idea intuitiva, planteada al inicio de esta sección, acerca de la independencia de los sucesos *obtener rey* y *obtener espadas*, tiene sentido, asumida la hipótesis de equiprobabilidad de forma más o menos consciente. ◁

El concepto de independencia puede extenderse a más de dos sucesos, procurando mantener una coherencia con la definición anterior. Así, si n sucesos, A_1, ..., A_n, son independientes, parece razonable que debiera serlo también un grupo cualquiera de ellos, por ejemplo, A_1, A_2 y A_3. Si sólo exigiéramos que

$$p(A_1 \cap \ldots \cap A_n) = p(A_1) \cdot \ldots \cdot p(A_n),$$

no garantizaríamos que, por ejemplo,

$$p(A_1 \cap A_2 \cap A_3) = p(A_1) \cdot p(A_2) \cdot p(A_3).$$

La definición deberá ser, entonces, más compleja.

Definición 3.3 *Sea (Ω, α, p) un espacio probabilístico y sean $A_1, \ldots, A_n$, n sucesos de α. Diremos que son mutuamente independientes si para cada subconjunto $A_{i_1}, \ldots, A_{i_k}$ de estos sucesos $(k = 2, 3, \ldots, n)$ se verifica*

$$p(A_{i_1} \cap \ldots \cap A_{i_k}) = p(A_{i_1}) \cdot \ldots \cdot p(A_{i_k}),$$

donde $i_1, \ldots, i_k$ varían entre 1 y n y son distintos entre sí.

Si la igualdad anterior se cumple para $k = 2$, es decir, si

$$p(A_{i_1} \cap A_{i_2}) = p(A_{i_1}) \cdot p(A_{i_2}),$$

diremos que los n sucesos son independientes dos a dos.

Obsérvese que n sucesos son mutuamente independientes, si la probabilidad de que se realicen simultáneamente varios de ellos, es igual al producto de sus probabilidades individuales; asimismo, son independientes dos a dos, si la probabilidad

de que se realicen simultáneamente dos cualesquiera de ellos, es igual al producto de sus probabilidades.

En particular, para tres sucesos tenemos la siguiente definición.

Definición 3.4 *Sea (Ω, α, p) un espacio probabilístico y A, B y C tres sucesos. Diremos que A, B y C son mutuamente independientes si*

$$\begin{aligned} p(A \cap B) &= p(A) \cdot p(B) \,, \\ p(A \cap C) &= p(A) \cdot p(C) \,, \\ p(B \cap C) &= p(B) \cdot p(C) \quad y \\ p(A \cap B \cap C) &= p(A) \cdot p(B) \cdot p(C) \,. \end{aligned}$$

Diremos que A, B y C son independientes dos a dos, si se cumplen las tres primeras igualdades.

Cuando n sucesos son mutuamente independientes diremos, sencillamente, que son independientes.

Nótese que, así como la definición de independencia de dos sucesos es bastante simple, no ocurre lo mismo cuando son más de dos. El motivo que subyace a este hecho es que parece razonable que si, por ejemplo, cuatro sucesos A, B, C y D son mutuamente independientes, también lo sean A y $B \cup C$, $A \cap C$ y $B \cup \overline{D}$, $A \cap B$, C y $\overline{D}$, etc. En otras palabras, resulta interesante, desde un punto de vista estadístico, y *lógico* desde un enfoque intuitivo, que, si varios sucesos son independientes, lo sean también sucesos formados a partir de los primeros, siempre que ninguno de éstos intervenga en más de uno de los nuevos sucesos. Puede el lector demostrar algunos resultados de este tipo.

Obviamente, si n sucesos son mutuamente independientes, son independientes dos a dos. Por el contrario, si son independientes dos a dos, no tienen por qué ser mutuamente independientes, como prueba el siguiente ejemplo, propuesto por S. N. BERNSTEIN.

Ejemplo 3.8. Se considera el espacio muestral $\Omega = \{1, 2, 3, 4\}$, finito y equiprobable. Demostraremos que los sucesos $A = \{1, 2\}$, $B = \{1, 3\}$ y $C = \{1, 4\}$ son independientes dos a dos, pero no mutuamente independientes.

Evidentemente, todas las posibles intersecciones de estos sucesos coinciden:

$$A \cap B = A \cap C = B \cap C = A \cap B \cap C = \{1\} \,.$$

Si calculamos sus probabilidades:

$$\begin{aligned} &p(A) = p(B) = p(C) = 1/2 \,, \\ &p(A \cap B) = p(A \cap C) = p(B \cap C) = 1/4 \quad y \\ &p(A \cap B \cap C) = 1/4 \,, \end{aligned}$$

deducimos que A, B y C son independientes dos a dos, pero no son mutuamente independientes. ◁

3.4 Composición de experimentos. Probabilidad en el espacio producto

Con frecuencia, consideraremos varios experimentos que se realizan de forma sucesiva, estando interesados en el resultado global de los mismos. Más en concreto, desearemos calcular la probabilidad de sucesos que tienen que ver con los diversos experimentos. Así, si lanzamos una moneda tres veces (cada lanzamiento constituye un experimento), querremos hallar, por ejemplo, la probabilidad de que *al menos se obtengan dos caras*. En este sentido, podemos hablar de un nuevo experimento, formado a partir de los tres experimentos simples: observar los resultados de los tres lanzamientos.

Es obvio que la situación descrita anteriormente también se puede entender considerando que existe un único experimento (resultado obtenido en los tres lanzamientos) que puede dividirse en fases o etapas (en este caso, tres fases, cada uno de los lanzamientos).

El cálculo de probabilidades de sucesos que tienen que ver con dos o más experimentos (lanzamiento de una moneda varias veces, extracción consecutiva de bolas de una urna, procesos de fabricación en etapas, selección de individuos de una población, etc.) se simplifica mucho si, por un lado, consideramos el experimento conjunto de los mismos y, por otro, calculamos las probabilidades ligadas a este experimento a partir de las correspondientes a los experimentos que lo integran.

Cuando realizamos los experimentos de forma sucesiva, las condiciones en que se realiza cada uno de ellos pueden estar influidas, o no, por el resultado de los experimentos precedentes. Así, cuando lanzamos la moneda tres veces, parece razonable suponer que la ocurrencia de *cara* en el primer lanzamiento no va a influir en el segundo, en el sentido de que las expectativas para ese segundo lanzamiento son las mismas salga *cara* o *cruz* en el primero (o no sepamos qué haya salido).

En este ejemplo, las tres fases en que se descompone el experimento (de hecho, tres experimentos aleatorios) son **físicamente independientes**, en el sentido de que los mecanismos que generan los resultados de cada uno de ellos no dependen del resultado que se haya obtenido en otro.

Esto no ocurre siempre. Otras veces, aunque el experimento aleatorio pueda dividirse en fases que a su vez sean experimentos aleatorios, dichas fases son **físicamente dependientes** en el sentido de que el resultado de una de dichas fases (o experimentos) modifica las condiciones en las que se desarrolla otra.

El siguiente ejemplo ilustra esta situación. Si tenemos una urna con dos bolas blancas y tres negras y realizamos sucesivas extracciones sin devolver la bola, parece razonable suponer que lo que ocurra en una extracción influye en la siguiente; evidentemente, si obtenemos una bola blanca en la primera extracción, la urna en la que haremos la segunda tendrá una bola blanca y tres negras, mientras que si obtenemos bola negra, nos encontramos en la segunda extracción ante una urna con dos bolas blancas y dos negras. La figura 3.3 nos describe esquemáticamente el proceso.

Por tanto, en esta situación, los experimentos se influyen mutuamente, presentan una dependencia física. El lector entenderá mejor esta circunstancia si se plantea la misma secuencia de extracciones cuando la bola observada se devuelve a

la urna. En este caso, al no cambiar su composición en los sucesivos experimentos, éstos no se influyen, son físicamente independientes.

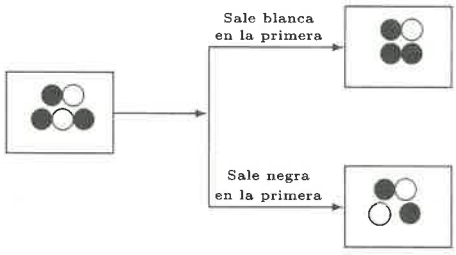


FIG. 3.3.

De cara al cálculo de las probabilidades de los sucesos relativos a varios experimentos, será necesario tener en cuenta si existe, o no, relación entre éstos, en el sentido descrito en los ejemplos anteriores. En general, cuando un experimento aleatorio se puede descomponer en varios (o, lo que es lo mismo, consideramos de forma conjunta varios experimentos), la asignación de probabilidades a los sucesos se facilita si aprovechamos, tanto la información sobre el mecanismo que relaciona las distintas fases, como las asignaciones de probabilidades que puedan realizarse en cada una de ellas. Además, procediendo de esta manera, los resultados obtenidos son más acordes con la intuición.

Las técnicas que ayudan a estudiar estos problemas reciben el nombre genérico de **composición de experimentos**, y van a ser desarrolladas en esta sección. El estudio se realiza para experimentos que se descomponen en dos, aunque las conclusiones se pueden extender, sin dificultad, a los que se componen de más de dos experimentos. Daremos la siguiente definición.

Definición 3.5 *Sean dos experimentos aleatorios que designaremos simbólicamente por E_1 y E_2. El experimento compuesto de E_1 y E_2 es el consistente en la realización de ambos de acuerdo con el mecanismo que los relaciona. Simbólicamente, si E es el experimento compuesto de E_1 y E_2, lo escribiremos como $E = E_1 \times E_2$.*

Los resultados del experimento compuesto tendrán que ver con los de cada uno de los experimentos que lo componen. La relación entre ellos se precisa en la siguiente definición.

Definición 3.6 *Sean E_1 y E_2 dos experimentos aleatorios y sean Ω_1 y Ω_2 los espacios muestrales asociados a E_1 y E_2, respectivamente. Se define el espacio muestral producto de Ω_1 y Ω_2, $\Omega_1 \times \Omega_2$, como el producto cartesiano de ambos, es decir,*

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega, \tau) \mid \omega \in \Omega_1, \tau \in \Omega_2\} .$$

El espacio muestral producto, $\Omega = \Omega_1 \times \Omega_2$, será el espacio muestral asociado al experimento aleatorio compuesto $E = E_1 \times E_2$.

Fijado el espacio muestral, otro problema que plantea la composición de experimentos es la selección de la σ-álgebra de sucesos en Ω; parece obvio que, en la determinación de esta σ-álgebra, se deberá tener en cuenta, de una u otra forma, las σ-álgebras de los experimentos E_1 y E_2. En concreto, si disponemos de una σ-álgebra, α_1, de sucesos de Ω_1, y otra σ-álgebra, α_2, de sucesos de Ω_2, el problema consiste en seleccionar una σ-álgebra, α, de sucesos de Ω, que sea congruente con α_1 y α_2. Por ejemplo, si A y B son sucesos de α_1 y α_2, respectivamente, parecería razonable que la ocurrencia de A en el primer experimento y de B en el segundo sea un suceso de Ω, es decir, que el subconjunto de Ω

$$A \times B = \{(\omega, \tau) \mid \omega \in A, \tau \in B\}$$

pertenezca a α.

La resolución del problema en toda su amplitud desborda el nivel de este libro, si bien se puede intuir que la σ-álgebra buscada es la mínima que contiene a $\alpha_1 \times \alpha_2$. Señalemos, no obstante, que si Ω_1 y Ω_2 son finitos o infinitos numerables y trabajamos, como es habitual en estos casos, con $\alpha_1 = \mathcal{P}(\Omega_1)$ y $\alpha_2 = \mathcal{P}(\Omega_2)$, tendremos que considerar en Ω la σ-álgebra formada por todos sus subconjuntos, esto es, $\alpha = \mathcal{P}(\Omega)$.

En la σ-álgebra α asociada al experimento compuesto, hay sucesos que dependen de los dos experimentos, mientras que otros sólo dependen de uno de ellos. Estos últimos pueden entenderse de dos formas: bien como sucesos de $\Omega = \Omega_1 \times \Omega_2$, bien como sucesos del espacio muestral correspondiente al experimento del que dependen. Esto nos lleva a la siguiente definición.

Definición 3.7 *Sea $E = E_1 \times E_2$ y sea $\Omega = \Omega_1 \times \Omega_2$ el espacio muestral producto asociado a E. Si A es un suceso de Ω_1, el suceso equivalente a A en Ω es el subconjunto de Ω*

$$A' = A \times \Omega_2 \, .$$

Si B es un suceso de Ω_2, el suceso equivalente a B en Ω es el subconjunto de Ω

$$B' = \Omega_1 \times B \, .$$

Además, diremos que A' es un suceso de Ω que depende del primer experimento y que B' es un suceso de Ω que depende del segundo experimento.

La anterior definición formaliza el hecho de que un suceso A del experimento E_1 se pueda describir tanto en Ω_1 como en Ω; en el espacio producto se describirá como *ocurre A en el primer experimento y cualquier resultado en el segundo.* Podemos hacer un comentario análogo para un suceso B de E_2.

Veamos cómo se concretan las anteriores definiciones en un ejemplo.

Ejemplo 3.9. Tenemos una urna con dos bolas blancas y tres negras y realizamos dos extracciones **con reemplazamiento**, observando el resultado obtenido. Para entender mejor el proceso diferenciaremos las dos etapas en que se puede descomponer el experimento.

EXPERIMENTO E_1: primera extracción.

El espacio muestral asociado a dicho experimento es

$$\Omega_1 = \{b_1, n_1\} ,$$

donde b_1 y n_1 son los dos posibles resultados de dicho experimento: obtener bola blanca y bola negra, respectivamente. Los subíndices, prescindibles, recalcarán el hecho de que se trata de resultados del primer experimento y, de esta forma, facilitarán las distinción con los correspondientes resultados del segundo experimento.

Como es habitual en este tipo de situaciones, utilizaremos la σ-álgebra

$$\mathcal{P}(\Omega_1) = \{\emptyset, \{b_1\}, \{n_1\}, \Omega_1\} .$$

Por tanto, $(\Omega_1, \mathcal{P}(\Omega_1))$ es el espacio probabilizable asociado al primer experimento.

EXPERIMENTO E_2: segunda extracción.

En este caso, el espacio muestral será

$$\Omega_2 = \{b_2, n_2\} ,$$

donde b_2 y n_2 representan los dos posibles resultados de la segunda extracción, sacar bola blanca y bola negra, respectivamente.

La σ-álgebra será:

$$\mathcal{P}(\Omega_2) = \{\emptyset, \{b_2\}, \{n_2\}, \Omega_2\} .$$

Por tanto, $(\Omega_2, \mathcal{P}(\Omega_2))$ es el espacio probabilizable asociado a E_2.

EXPERIMENTO COMPUESTO $E = E_1 \times E_2$: primera y segunda extracción.

El espacio muestral será el producto cartesiano de los espacios muestrales de los experimentos E_1 y E_2. Es decir,

$$\Omega = \Omega_1 \times \Omega_2 = \{(b_1, b_2), (b_1, n_2), (n_1, b_2), (n_1, n_2)\} ,$$

donde (b_1, b_2) representa el resultado de obtener bola blanca en la primera y segunda extracción, (b_1, n_2) blanca en la primera y negra en la segunda, etc.

La σ-álgebra sobre el espacio producto será $\mathcal{P}(\Omega)$. Esta σ-álgebra contendrá 16 sucesos, incluidos $\emptyset$ y Ω.

Entonces, $(\Omega = \Omega_1 \times \Omega_2, \mathcal{P}(\Omega))$ será el espacio probabilizable asociado al experimento compuesto.

Si realizamos las dos extracciones **sin reemplazamiento** no cambiará la descomposición del experimento en dos, así como los espacios probabilizables asociados a E_1, E_2 y $E = E_1 \times E_2$.

Las diferencias entre las dos extracciones con devolución (experimentos físicamente independientes) y las dos extracciones sin devolución (experimentos físicamente dependientes) no están, por tanto, en sus correspondientes espacios probabilizables asociados, sino, como veremos más adelante, en los espacios probabilísticos, es decir, en las probabilidades que se asignarán en uno y otro caso.

Advertirá el lector que hay sucesos en el experimento compuesto E que dependerán de los dos experimentos que lo forman, E_1 y E_2. Así, el suceso *obtener al menos una bola blanca*, es decir,

$$\{(b_1, b_2), (b_1, n_2), (n_1, b_2)\} ,$$

depende, tanto de la primera extracción, como de la segunda.

En cambio, hay sucesos en E que sólo dependen de uno de los experimentos que lo componen. Así, *obtener bola blanca en la primera extracción*, es decir,

$$\{(b_1, b_2), (b_1, n_2)\} ,$$

sólo depende del primer experimento E_1. Dado que $\Omega_2 = \{b_2, n_2\}$, dicho suceso lo podemos escribir como

$$\{b_1\} \times \Omega_2 \ ,$$

describiéndose literalmente como *obtener bola blanca en la primera extracción y cualquiera en la segunda.*

Este tipo de sucesos se puede expresar en términos del experimento del que dependen. Así, si consideramos *obtener bola blanca en la primera extracción* como suceso del experimento E_1, éste será:

$$\{b_1\} \ .$$

Los sucesos

$$\{(b_1, b_2), (b_1, n_2)\} = \{b_1\} \times \Omega_2$$

y

$$\{b_1\}$$

son el mismo suceso, formalizado en distintos espacios muestrales; por un lado, en el experimento compuesto $E = E_1 \times E_2$, y, por otro, en el experimento E_1. Se trata, por tanto, de sucesos equivalentes.

Análogamente, podemos describir el suceso *obtener negra en la segunda extracción* como suceso de E,

$$\{(b_1, n_2), (n_1, n_2)\} = \Omega_1 \times \{n_2\} \ ,$$

o como suceso de E_2,

$$\{n_2\}. \triangleleft$$

3.4.1 Asignación de probabilidades en experimentos compuestos

Otro problema es el que se refiere a la asignación de probabilidades a los sucesos del espacio muestral producto. Si un experimento E está compuesto de dos, E_1 y E_2, con espacios muestrales Ω_1 y Ω_2 y σ-álgebras α_1 y α_2, acabamos de indicar cómo construir el espacio probabilizable ligado a E, (Ω, α), donde $\Omega = \Omega_1 \times \Omega_2$ y α se obtiene a partir de α_1 y α_2. Veamos cómo definir una probabilidad, p, sobre α, que sea congruente con la información de tipo probabilístico que nos suministren E_1, E_2 y la relación entre ambos. El objetivo de este apartado es, por tanto, estudiar cómo dicha información puede servirnos para asociar una estructura probabilística al experimento compuesto.

Estudiaremos esta cuestión suponiendo que Ω_1 y Ω_2 son infinito numerables y que α_1 y α_2 son, respectivamente, $\mathcal{P}(\Omega_1)$ y $\mathcal{P}(\Omega_2)$. Aunque los resultados puedan generalizarse a todo tipo de espacios muestrales y σ-álgebras, el desarrollo bajo supuestos más amplios es complejo.

Si escribimos

$$\Omega_1 = \{\omega_1, \ldots, \omega_n, \ldots\}$$

y

$$\Omega_2 = \{\tau_1, \ldots, \tau_m, \ldots\} \ ,$$

el espacio muestral producto, $\Omega = \Omega_1 \times \Omega_2$, será

$$\Omega = \{(\omega_i, \tau_j), i = 1, 2, \ldots; j = 1, 2, \ldots\} \ .$$

Como comentamos anteriormente, si $\alpha_1 = \mathcal{P}(\Omega_1)$ y $\alpha_2 = \mathcal{P}(\Omega_2)$, entonces $\alpha = \mathcal{P}(\Omega)$, esto es, todos los subconjuntos de Ω son sucesos. Sabemos que, en esta situación, queda definida una probabilidad, p, sobre α, asignando a cada suceso elemental (ω_i, τ_j) de Ω un número $p[\omega_i, \tau_j]$, de forma que:

1. $p[\omega_i, \tau_j] \geq 0$, $\quad i = 1, 2, \ldots ; j = 1, 2, \ldots$
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p[\omega_i, \tau_j] = 1$,

calculándose las probabilidades de cualquier suceso como suma de las probabilidades de los sucesos elementales que lo forman.

Este objetivo es fácilmente alcanzable en ciertas situaciones: aquellas en las que puede suponerse que los dos experimentos son independientes, es decir, que el resultado de uno de ellos no influirá sobre el otro. Cuando esto ocurre, definiremos p de la siguiente forma.

Definición 3.8 *Sea $E = E_1 \times E_2$ un experimento compuesto, y $(\Omega_1, \alpha_1, p_1)$ y $(\Omega_2, \alpha_2, p_2)$ los espacios probabilísticos asociados a E_1 y E_2, respectivamente, con Ω_1 y Ω_2 infinito numerables, $\alpha_1 = \mathcal{P}(\Omega_1)$ y $\alpha_2 = \mathcal{P}(\Omega_2)$. Sea (Ω, α) el espacio probabilizable asociado a E, con $\Omega = \Omega_1 \times \Omega_2$ y $\alpha = \mathcal{P}(\Omega)$. Si los experimentos E_1 y E_2 son independientes, la probabilidad, p, sobre α, se define como aquella que verifica*

$$p[\omega_i, \tau_j] = p_1(\omega_i) \cdot p_2(\tau_j) \ , \qquad i = 1, 2, \ldots ; j = 1, 2, \ldots$$

Comprobemos que la probabilidad que acabamos de definir lo es realmente. Teniendo en cuenta que dicha asignación se ha realizado sobre un espacio muestral Ω infinito numerable y que la σ-álgebra es $\mathcal{P}(\Omega)$, nos bastará con demostrar que:

1. $p[\omega_i, \tau_j] \geq 0$.
2. $\sum_{(\omega_i, \tau_j) \in \Omega} p[\omega_i, \tau_j] = 1$.

Demostración. Evidentemente, $p[\omega_i, \tau_j]$ es no negativa por ser el producto de dos números no negativos. Además,

$$\begin{aligned} p(\Omega) = \sum_{(\omega_i, \tau_j) \in \Omega} p[\omega_i, \tau_j] &= \sum_{\omega_i \in \Omega_1} \sum_{\tau_j \in \Omega_2} p_1(\omega_i) \cdot p_2(\tau_j) = \\ &= \sum_{\omega_i \in \Omega_1} p_1(\omega_i) \cdot \sum_{\tau_j \in \Omega_2} p_2(\tau_j) = 1. \clubsuit \end{aligned}$$

En definitiva, (Ω, α, p), así definido, es un espacio probabilístico asociado al experimento compuesto. Como es conocido, si $S \subset \Omega$, su probabilidad se calcula como

$$p(S) = \sum_{(\omega_i, \tau_j) \in S} p[\omega_i, \tau_j] \ .$$

Pero, además, según veremos en las dos proposiciones siguientes, esta probabilidad p, así definida, es la más coherente que podemos construir sobre Ω en un doble

sentido: en primer lugar, *respeta* las probabilidades de los experimentos E_1 y E_2, y, en segundo lugar, conecta la idea intuitiva de independencia de experimentos con la definición formal de independencia de sucesos.

La primera de estas proposiciones nos va a confirmar que la probabilidad de un suceso, que tenga que ver con un único experimento, se puede calcular indistintamente en el experimento compuesto E, es decir, con la probabilidad p definida en 3.8 o en el experimento del que dependen (E_1 o E_2), es decir, con la probabilidad (p_1 o p_2) definida en el mismo.

Proposición 3.8 *Sea $E = E_1 \times E_2$ un experimento compuesto, y $(\Omega_1, \alpha_1, p_1)$ y $(\Omega_2, \alpha_2, p_2)$ los espacios probabilísticos asociados a los experimentos independientes E_1 y E_2. Sea (Ω, α, p) el espacio probabilístico asociado a E, donde p está construida de acuerdo con la definición 3.8. Entonces, si A es un suceso de Ω_1 y A' es el suceso equivalente a A en Ω,*

$$p(A') = p_1(A) .$$

Igualmente, si B es un suceso de Ω_2 y B' es el suceso equivalente a B en Ω,

$$p(B') = p_2(B) .$$

Demostración. Si los experimentos son independientes y A' sólo depende del primer experimento, es decir, $A' = A \times \Omega_2$, se tiene que

$$\begin{aligned} p(A') &= \sum_{(\omega_i,\tau_j)\in A'} p[\omega_i, \tau_j] = \sum_{\omega_i \in A} \sum_{\tau_j \in \Omega_2} p_1(\omega_i) \cdot p_2(\tau_j) = \\ &= \sum_{\omega_i \in A} p_1(\omega_i) \cdot \sum_{\tau_j \in \Omega_2} p_2(\tau_j) = \sum_{\omega_i \in A} p_1(\omega_i) = p_1(A) . \end{aligned}$$

Análogamente, se probaría el resultado para el suceso dependiente del segundo experimento. ♣

Estas probabilidades, p_1 y p_2, reciben el nombre de **probabilidades marginales** de la probabilidad p.

A continuación, demostraremos una nueva proposición que relaciona la independencia de sucesos con la independencia de experimentos, afirmando, de hecho, que un suceso que sólo depende de E_1 y otro que sólo depende de E_2 son independientes si lo son los dos experimentos.

Proposición 3.9 *Sean E_1 y E_2 dos experimentos independientes y (Ω, α, p) el espacio probabilístico asociado al experimento compuesto $E = E_1 \times E_2$. Entonces, si A es un suceso de Ω_1 y B un suceso de Ω_2, los sucesos $A' = A \times \Omega_2$ y $B' = \Omega_1 \times B$ son estadísticamente independientes, esto es,*

$$p(A' \cap B') = p(A') \cdot p(B') .$$

Demostración. En efecto,

$$A' \cap B' = (A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B .$$

Entonces,

$$\begin{aligned} p(A' \cap B') = p(A \times B) &= \sum_{\omega_i \in A} \sum_{\tau_j \in B} p[\omega_i, \tau_j] = \sum_{\omega_i \in A} \sum_{\tau_j \in B} p_1(\omega_i) \cdot p_2(\tau_j) = \\ &= \sum_{\omega_i \in A} p_1(\omega_i) \cdot \sum_{\tau_j \in B} p_2(\tau_j) = p_1(A) \cdot p_2(B) = \\ &= p(A') \cdot p(B') \,. \clubsuit \end{aligned}$$

El lector podría demostrar que la probabilidad p, definida en 3.8, es la única probabilidad consistente en el doble sentido que ponen de manifiesto las dos proposiciones anteriores.

Ejemplo 3.10. Consideramos de nuevo el experimento de realizar dos extracciones **con reemplazamiento** de una urna que contiene dos bolas blancas y tres negras. Hemos descrito los espacios probabilizables asociados, tanto al experimento compuesto, como a cada uno de los dos que lo componen. Calculemos, ahora, las probabilidades en cada uno de los experimentos.

EXPERIMENTO E_1: primera extracción.

Dada la naturaleza del experimento, podemos suponer que cada uno de las cinco bolas de la urna puede ser elegida con la misma probabilidad, 1/5. Entonces,

$$p_1(b_1) = \frac{2}{5}$$

y

$$p_1(n_1) = \frac{3}{5} \,,$$

con lo cual, tenemos perfectamente determinado $(\Omega_1, \mathcal{P}(\Omega_1), p_1)$, espacio probabilístico asociado al primer experimento.

EXPERIMENTO E_2: segunda extracción.

Dado que devolvemos la bola obtenida en la primera extracción, realizaremos la segunda bajo las mismas condiciones. Por tanto,

$$p_2(b_2) = \frac{2}{5}$$

y

$$p_2(n_2) = \frac{3}{5} \,.$$

Así, queda determinado el espacio probabilístico $(\Omega_2, \mathcal{P}(\Omega_2), p_2)$ asociado al segundo experimento.

EXPERIMENTO COMPUESTO $E = E_1 \times E_2$: primera y segunda extracción.

Asignaremos una probabilidad a los cuatro sucesos elementales de E, teniendo en cuenta que los experimentos son físicamente independientes, es decir, que lo que ocurre en la primera extracción no influye en la segunda. Entonces,

$$\begin{aligned} p[b_1, b_2] = p_1(b_1) \cdot p_2(b_2) &= \frac{2}{5} \cdot \frac{2}{5} = \frac{4}{25} \,, \\ p[b_1, n_2] = p_1(b_1) \cdot p_2(n_2) &= \frac{2}{5} \cdot \frac{3}{5} = \frac{6}{25} \,, \\ p[n_1, b_2] = p_1(n_1) \cdot p_2(b_2) &= \frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25} \quad \text{y} \\ p[n_1, n_2] = p_1(n_1) \cdot p_2(n_2) &= \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{25} \,. \end{aligned}$$

Puede comprobar el lector que la asignación es correcta, es decir, son valores positivos que suman uno.

También puede adivinarse que el resultado obtenido está de acuerdo con la idea intuitiva de independencia. Así, si se espera obtener bola blanca en la primera extracción en el 40 % de las ocasiones ($p_1(b_1) = 0.4$) y bola negra en la segunda extracción en el 60 % de las ocasiones ($p_2(n_2) = 0.6$), y, además, el hecho de obtener bola blanca en la primera no influye para nada en la segunda extracción, entonces esperamos obtener bola blanca en la primera y negra en la segunda en el 40 % del 60 % de las ocasiones, es decir, en el 24 %. En términos probabilísticos,

$$p_1(b_1) \cdot p_2(n_2) = 0.4 \cdot 0.6 = 0.24 = p[b_1, n_2] \, .$$

Una vez asignada la probabilidad a los sucesos elementales, queda determinada la probabilidad de cualquier otro suceso. Por ejemplo, para el suceso *obtener al menos una bola blanca*,

$$\begin{aligned} p(\text{obtener al menos una bola blanca}) &= p[(b_1, b_2), (b_1, n_2), (n_1, b_2)] = \\ &= p[b_1, b_2] + p[b_1, n_2] + p[n_1, b_2] = \\ &= \frac{4}{25} + \frac{6}{25} + \frac{6}{25} = \frac{16}{25} \, , \end{aligned}$$

o para el suceso *obtener bola blanca en la primera extracción*,

$$p(\text{obtener bola blanca en la primera extracción}) = p[(b_1, b_2), (b_1, n_2)] = \frac{10}{25} \, .$$

Adviértase que este suceso sólo depende del primer experimento, con lo cual podemos calcular en dicho experimento su probabilidad, es decir,

$$p_1(b_1) = \frac{2}{5} \, .$$

Como hemos probado anteriormente, al tratarse del mismo suceso formalizado en distintos espacios muestrales, las probabilidades tienen que coincidir:

$$p[(b_1, b_2), (b_1, n_2)] = p_1(b_1) \, .$$

Asimismo, puede el lector considerar el suceso *obtener bola negra en la segunda extracción*, calculando su probabilidad, bien en el experimento compuesto E, o bien en el experimento E_2. Evidentemente, ambas probabilidades coinciden, siendo su valor 3/5.

Se comprueba, además, que los sucesos *obtener bola blanca en la primera extracción* y *obtener bola negra en la segunda* son independientes en el experimento compuesto, independencia que está motivada por el hecho de que cada uno de estos sucesos sólo depende de uno de los experimentos independientes que forman el experimento compuesto. Así,

$$\begin{aligned} p[\{(b_1, b_2), (b_1, n_2)\} \cap \{(b_1, n_2), (n_1, n_2)\}] &= p[b_1, n_2] = \frac{6}{25} = \\ &= p[(b_1, b_2), (b_1, n_2)] \cdot p[(b_1, n_2), (n_1, n_2)] \, . \triangleleft \end{aligned}$$

Cuando los experimentos son físicamente dependientes, la construcción de una probabilidad no puede realizarse con la técnica anterior, porque conduciría a resultados contrarios a la intuición (la probabilidad de un suceso en uno cualquiera de los experimentos no dependería de los resultados obtenidos en el otro). Así,

podríamos definir experimentos dependientes como aquellos que no son independientes, es decir, aquellos a los que no tiene sentido asignar conjuntamente una probabilidad como la definida en 3.8.

No existe una técnica universal que permita asignar una probabilidad razonable a un experimento compuesto de experimentos dependientes, ya que su construcción estará determinada por el tipo de dependencia que exista. Con todo, muchas veces los experimentos se desarrollan sucesivamente en el tiempo, y la relación tiene un sentido unidireccional (el resultado del primer experimento modifica las condiciones en las que se realiza posteriormente el segundo). En estas situaciones, se conocerá la probabilidad p_1 asociada a Ω_1, y una colección de probabilidades sobre Ω_2, $\{p_{\omega_i},\ \omega_i \in \Omega_1\}$, donde $p_{\omega_i}(\tau_j)$ sería la probabilidad de obtener τ_j en el segundo experimento cuando el resultado del primero ha sido ω_i. Estas probabilidades se denominan **probabilidades de transición**, porque permiten pasar del primer al segundo experimento. Aunque la situación que estamos estudiando es muy particular, conviene fijar, mediante una definición, cómo se construye la probabilidad más adecuada en este contexto.

Definición 3.9 *Sea $E = E_1 \times E_2$ un experimento compuesto. Sea $(\Omega_1, \alpha_1, p_1)$ el espacio probabilístico asociado a E_1 y, para cada $\omega_i \in \Omega_1$, sea $(\Omega_2, \alpha_2, p_{\omega_i})$ el espacio probabilístico asociado a E_2 conocido el resultado, ω_i, del experimento E_1. Se define, entonces, una probabilidad, p, en $\Omega = \Omega_1 \times \Omega_2$ como*

$$p[(\omega_i, \tau_j)] = p_1(\omega_i) \cdot p_{\omega_i}(\tau_j)\ ,$$

donde (ω_i, τ_j) es un resultado cualquiera de Ω.

La aplicación, así construida, es realmente una probabilidad. En efecto,

1. $p[(\omega_i, \tau_j)] \geq 0$, por ser el producto de dos números no negativos.
2. La suma de las probabilidades de los sucesos elementales de Ω es la unidad; en efecto:

$$\begin{aligned}
\sum_{(\omega_i,\tau_j)\in\Omega} p[(\omega_i, \tau_j)] &= \sum_{\omega_i\in\Omega_1} \sum_{\tau_j\in\Omega_2} p_1(\omega_i) \cdot p_{\omega_i}(\tau_j) = \\
&= \sum_{\omega_i\in\Omega_1} \left(p_1(\omega_i) \cdot \sum_{\tau_j\in\Omega_2} p_{\omega_i}(\tau_j) \right) = \\
&= \sum_{\omega_i\in\Omega_1} (p_1(\omega_i) \cdot 1) = 1\ ,
\end{aligned}$$

ya que, tanto p_1 como p_{ω_i}, para cualquier ω_i, son probabilidades.

Esta definición es coherente con la información probabilística que se tiene sobre los experimentos y la relación entre los mismos. Como veremos a continuación, la probabilidad de un suceso que sólo dependa del primer experimento puede calcularse, tanto en dicho experimento (con p_1), como en el experimento compuesto (con

p); además, la probabilidad de obtener el resultado τ_j en el segundo experimento, condicionada por haber obtenido ω_i en el primero, coincide con la probabilidad de transición $p_{\omega_i}(\tau_j)$ definida sobre Ω_2.

Proposición 3.10 *Sea $E = E_1 \times E_2$ un experimento compuesto y $(\Omega_1, \alpha_1, p_1)$ el espacio probabilístico asociado a E_1. Sea (Ω, α, p) el espacio probabilístico asociado a E, donde p está construida de acuerdo con la definición 3.9. Entonces:*

1. *Si A es un suceso de Ω_1 y $A' = A \times \Omega_2$ es el suceso equivalente a A en Ω,*

$$p(A') = p_1(A) .$$

2. *La probabilidad de obtener τ_j en el segundo experimento condicionada por haber obtenido ω_i en el primero, abreviadamente $p(\tau_j \mid \omega_i)$, coincide con la probabilidad de transición $p_{\omega_i}(\tau_j)$.*

Demostración. En primer lugar,

$$p(A') = p(A \times \Omega_2) = \sum_{(\omega_i, \tau_j) \in A \times \Omega_2} p[(\omega_i, \tau_j)] =$$

$$= \sum_{\omega_i \in A} \left(p_1(\omega_i) \cdot \sum_{\tau_j \in \Omega_2} p_{\omega_i}(\tau_j) \right) = p_1(A) ,$$

ya que, como las probabilidades de transición lo son,

$$\sum_{\tau_j \in \Omega_2} p_{\omega_i}(\tau_j) = 1 ,$$

sea cual sea $\omega_i \in \Omega_1$.

En segundo lugar,

$$p(\tau_j \mid \omega_i) = p(\Omega_1 \times \{\tau_j\} \mid \{\omega_i\} \times \Omega_2) = \frac{p[(\omega_i, \tau_j)]}{p(\omega_i)} ,$$

donde en el numerador aparece la probabilidad de la intersección de los dos sucesos, $\Omega_1 \times \{\tau_j\}$ y $\{\omega_i\} \times \Omega_2$, y en el denominador, de forma abreviada, la probabilidad del segundo. Pero, $p[(\omega_i, \tau_j)] = p_1(\omega_i) \cdot p_{\omega_i}(\tau_j)$ y, por el resultado anterior, $p(\omega_i) = p_1(\omega_i)$. En consecuencia,

$$p(\tau_j \mid \omega_i) = \frac{p_1(\omega_i) \cdot p_{\omega_i}(\tau_j)}{p_1(\omega_i)} = p_{\omega_i}(\tau_j) . \clubsuit$$

La proposición anterior demuestra que la asignación de probabilidad en Ω, realizada en la definición 3.9, es coherente con lo que se conoce sobre el experimento compuesto. En efecto, no cambia la probabilidad de sucesos que dependen sólo del primer experimento (probabilidad p_1 que ya es conocida); además, las probabilidades de los resultados del segundo experimento, condicionadas por el resultado obtenido en el primero, muestran la forma en que aquél depende de éste, y coinciden, por tanto, con las probabilidades de transición. El lector puede comprobar

que la probabilidad dada por la definición 3.9, es la única probabilidad consistente en el sentido de la proposición 3.10. Por lo que se refiere a las probabilidades de sucesos que sólo dependen del segundo experimento, obsérvese que no son conocidas, por lo que no hay coherencia exigible. De hecho, las probabilidades de dichos sucesos sólo pueden calcularse en el experimento compuesto. En el fondo, se define una probabilidad sobre Ω_2 de la siguiente forma.

Definición 3.10 *Sea $E = E_1 \times E_2$ un experimento compuesto y sea (Ω, α, p) el espacio probabilístico asociado a E, donde p está construida de acuerdo con la definición 3.9. Definimos una probabilidad p_2 en Ω_2 como*

$$p_2(B) = p(B') \quad \textit{para todo } B \subset \Omega_2,$$

donde B' es el suceso equivalente a B en Ω, esto es, $B' = \Omega_1 \times B$.

Así, hemos determinado el espacio probabilístico $(\Omega_2, \alpha_2, p_2)$ asociado al segundo experimento.

De nuevo, p_1 y p_2 reciben el nombre de **probabilidades marginales** de la probabilidad p. Adviértase, por otro lado, que, al contrario de lo que ocurría para experimentos independientes, el conocimiento de p_1 y p_2 no es suficiente para determinar p. Dicho de otra forma, idénticas probabilidades marginales se pueden corresponder con distintas probabilidades en el espacio producto.

Ejemplo 3.11. Volvamos a considerar el experimento de realizar dos extracciones **sin reemplazamiento** de una urna que contiene dos bolas blancas y tres negras. Anteriormente hemos descrito los espacios probabilizables asociados a cada uno de los experimentos. Hagamos ahora las correspondientes asignaciones de probabilidad.

EXPERIMENTO E_1: primera extracción.

En esta fase inicial no influye la relación entre la primera y segunda extracción; es decir, el hecho de que no devolvamos la bola obtenida en la primera, cambia la composición de la urna para la segunda extracción, pero no influye en la determinación del espacio probabilístico en esta fase inicial. Entonces,

$$p_1(b_1) = \frac{2}{5}$$

y

$$p_1(n_1) = \frac{3}{5}.$$

De esta forma queda determinado $(\Omega_1, \mathcal{P}(\Omega_1), p_1)$, espacio probabilístico asociado a E_1.

A continuación, parecería lógico que tratásemos de determinar la probabilidad del experimento E_2. Ahora bien, a la hora de asignar esta probabilidad, debemos tener en cuenta las condiciones en que se realiza este experimento, condiciones que vienen determinadas por lo ocurrido en el primer experimento. Es decir, tendremos que considerar los dos experimentos de forma conjunta. Por tanto, abordaremos, previamente, la asignación de probabilidad en el experimento compuesto.

EXPERIMENTO COMPUESTO $E = E_1 \times E_2$: primera y segunda extracción.

Al igual que cuando las extracciones se realizaban con devolución, intentaremos determinar las probabilidades de los cuatro sucesos elementales. Previamente, tendremos en cuenta las probabilidades de transición, es decir, las probabilidades de la segunda extracción dependiendo de lo ocurrido en la primera. La naturaleza del experimento y la

secuencia temporal del mismo, permiten calcular estas probabilidades de forma sencilla. Así, cuando sale bola blanca en la primera extracción, la urna en la que se realiza la segunda está compuesta por una bola blanca y tres negras. Entonces,

$$p_{b_1}(b_2) = \frac{1}{4}$$

y

$$p_{b_1}(n_2) = \frac{3}{4} .$$

Lógicamente esta probabilidad de transición está definida sobre $(\Omega_2, \mathcal{P}(\Omega_2))$, es decir, constituye el espacio probabilístico $(\Omega_2, \mathcal{P}(\Omega_2), p_{b_1})$.

Si sale bola negra en la primera extracción, la probabilidad de transición será:

$$p_{n_1}(b_2) = \frac{2}{4}$$

y

$$p_{n_1}(n_2) = \frac{2}{4} ,$$

concretándose el espacio probabilístico $(\Omega_2, \mathcal{P}(\Omega_2), p_{n_1})$.

Determinadas estas probabilidades, podemos calcular ya las de los sucesos del experimento compuesto:

$$\begin{aligned}
p[(b_1, b_2)] = p_1(b_1) \cdot p_{b_1}(b_2) &= \frac{2}{5} \cdot \frac{1}{4} = \frac{2}{20} , \\
p[(b_1, n_2)] = p_1(b_1) \cdot p_{b_1}(n_2) &= \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} , \\
p[(n_1, b_2)] = p_1(n_1) \cdot p_{n_1}(b_2) &= \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} \quad \text{y} \\
p[(n_1, n_2)] = p_1(n_1) \cdot p_{n_1}(n_2) &= \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} .
\end{aligned}$$

Esta asignación a los sucesos elementales nos permite calcular la probabilidad de cualquier otro suceso; por ejemplo, la de *obtener al menos una bola blanca*,

$$\begin{aligned}
p(\text{obtener al menos una bola blanca}) &= p[(b_1, b_2), (b_1, n_2), (n_1, b_2)] = \\
&= p[(b_1, b_2)] + p[(b_1, n_2)] + p[(n_1, b_2)] = \\
&= \frac{2}{20} + \frac{6}{20} + \frac{6}{20} = \frac{14}{20} ,
\end{aligned}$$

o la de *obtener bola blanca en la primera extracción*,

$$p(\text{obtener bola blanca en la primera extracción}) = p[(b_1, b_2), (b_1, n_2)] = \frac{8}{20} .$$

Dado que este último suceso sólo depende del primer experimento, su probabilidad se puede calcular en E_1:

$$p_1(b_1) = \frac{2}{5} .$$

Evidentemente, estas dos probabilidades coinciden.

Si consideramos un suceso cuya probabilidad sólo depende del segundo experimento, por ejemplo, *obtener bola negra en la segunda extracción*,

$$p(\text{obtener bola negra en la segunda extracción}) = p[(b_1, n_2), (n_1, n_2)] = \frac{12}{20} .$$

En principio, aunque este suceso sólo depende del segundo experimento, no podemos hallar su probabilidad en él pues aún no la hemos definido.

El lector puede comprobar que los sucesos *obtener bola blanca en la primera extracción* y *obtener bola negra en la segunda extracción* no son independientes en el experimento compuesto.

Consideremos los sucesos que sólo dependen de E_1,

$$\{b_1\} \times \Omega_2 = \{(b_1, b_2), (b_1, n_2)\}$$

y

$$\{n_1\} \times \Omega_2 = \{(n_1, b_2), (n_1, n_2)\} ,$$

y aquellos que sólo dependen de E_2,

$$\Omega_1 \times \{b_2\} = \{(b_1, b_2), (n_1, b_2)\}$$

y

$$\Omega_1 \times \{n_2\} = \{(b_1, n_2), (n_1, n_2)\} .$$

Calculemos en el experimento compuesto las probabilidades de los sucesos que sólo dependen del segundo experimento condicionadas por los que sólo dependen del primero. Por ejemplo, la probabilidad de *obtener bola blanca en la segunda extracción*, habiendo obtenido *blanca en la primera* será:

$$\begin{aligned} p[\Omega_1 \times \{b_2\} \mid \{b_1\} \times \Omega_2] &= \frac{p[(\Omega_1 \times \{b_2\}) \cap (\{b_1\} \times \Omega_2)]}{p[\{b_1\} \times \Omega_2]} = \frac{p[(b_1, b_2)]}{p[(b_1, b_2), (b_1, n_2)]} = \\ &= \frac{2/20}{8/20} = \frac{1}{4} = p_{b_1}(b_2) . \end{aligned}$$

Como se ve, esta probabilidad coincide con una probabilidad de transición, en concreto, con la probabilidad de *obtener bola blanca en la segunda extracción* cuando se obtuvo bola blanca en la primera. Igualmente, puede comprobarse que

$$\begin{aligned} p[\Omega_1 \times \{n_2\} \mid \{b_1\} \times \Omega_2] &= p_{b_1}(n_2) , \\ p[\Omega_1 \times \{b_2\} \mid \{n_1\} \times \Omega_2] &= p_{n_1}(b_2) \quad \text{y} \\ p[\Omega_1 \times \{n_2\} \mid \{n_1\} \times \Omega_2] &= p_{n_1}(n_2) , \end{aligned}$$

resultados, por otro lado, absolutamente razonables.

Todo lo anterior nos permite simplificar la notación, siempre que ello no induzca a confusión. Así, la probabilidad de *obtener bola blanca en la primera extracción y negra en la segunda* la podemos escribir como

$$p[(b_1, n_2)] = p(b_1) \cdot p(n_2 \mid b_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} ,$$

que no es otra cosa que aplicar la regla de la multiplicación en el experimento compuesto, simplificando las notaciones de sucesos y probabilidades. Advierta el lector que utilizando estrictamente la notación, deberíamos escribir

$$\begin{aligned} p[(b_1, n_2)] &= p[(\{b_1\} \times \Omega_2) \cap (\Omega_1 \times \{n_2\})] = \\ &= p[\{b_1\} \times \Omega_2] \cdot p[\Omega_1 \times \{n_2\} \mid \{b_1\} \times \Omega_2] = \\ &= p_1(b_1) \cdot p_{b_1}(n_2) = \frac{2}{5} \cdot \frac{3}{4} = \frac{6}{20} . \end{aligned}$$

La comparación de las dos expresiones ilustra muy a las claras la necesidad de simplificar la notación.

Una vez determinado el espacio probabilístico, $(\Omega, \mathcal{P}(\Omega), p)$, en el experimento compuesto, podemos calcular las probabilidades en el segundo experimento.

EXPERIMENTO E_2: segunda extracción.

La coherencia que debe exigirse a estas asignaciones de probabilidad obliga a definir la probabilidad de un suceso en E_2 a partir de la probabilidad del mismo suceso en E. De esta forma,

$$p_2(b_2) = p[(b_1, b_2), (n_1, b_2)] = \frac{8}{20}$$

y

$$p_2(n_2) = p[(b_1, n_2), (n_1, n_2)] = \frac{12}{20}\,.$$

Téngase en cuenta que esta asignación de probabilidad coincide con la realizada en el primer experimento. Reflexione el lector sobre esta coincidencia aparentemente paradójica, pero perfectamente razonable.

De esta forma determinamos el espacio probabilístico asociado al segundo experimento, $(\Omega_2, \mathcal{P}(\Omega_2), p_2)$.

Adviértase que las probabilidades en el experimento compuesto son distintas si las extracciones se hacen con o sin reemplazamiento; por el contrario, las marginales coinciden.

Una vez hechas todas las consideraciones anteriores, estamos en disposición de afrontar cálculos de probabilidades condicionadas siguiendo la secuencia inversa a aquella en la que se realizan los experimentos. Estos cálculos son habituales cuando el primer experimento selecciona una situación (una causa) entre varias posibles y el segundo proporciona un resultado (un efecto) dada una situación de partida. Con frecuencia se conoce el efecto (el resultado del segundo experimento), pero no la causa que lo provocó (el resultado del primero). Por ejemplo, sabemos que se ha obtenido bola negra en la segunda extracción, ¿cuál será la probabilidad de que en la primera saliese blanca? Escribiéndolo con una notación rigurosa,

$$\begin{aligned} p[\{b_1\} \times \Omega_2 \mid \Omega_1 \times \{n_2\}] &= \frac{p[(\{b_1\} \times \Omega_2) \cap (\Omega_1 \times \{n_2\})]}{p[\Omega_1 \times \{n_2\}]} = \\ &= \frac{p[(b_1, n_2)]}{p[(b_1, n_2), (n_1, n_2)]} = \frac{1}{2}\,. \end{aligned}$$

El resultado anterior es una aplicación del teorema de Bayes para sucesos del experimento compuesto. Así, en notación abreviada,

$$p(b_1 \mid n_2) = \frac{p[(b_1, n_2)]}{p(n_2)} = \frac{p(b_1) \cdot p(n_2 \mid b_1)}{p(b_1) \cdot p(n_2 \mid b_1) + p(n_1) \cdot p(n_2 \mid n_1)} = \frac{1}{2}\,. \triangleleft$$

Tres resultados importantes de este capítulo, regla de la multiplicación, teorema de la probabilidad total y teorema de Bayes, alcanzan ahora todas sus virtualidades en el sentido del ejemplo anterior; téngase en cuenta que la mayor parte de las aplicaciones de estos resultados se presentan en situaciones de experimentos compuestos, o lo que es lo mismo, experimentos que se realizan en fases sucesivas.

Por último, recordemos de nuevo al lector que todos los resultados demostrados en esta sección son aplicables en un contexto más amplio: experimentos que se descomponen en más de dos fases y que tienen asociados espacios probabilizables cualesquiera.

3.4.2 Equiprobabilidad en experimentos compuestos

Las técnicas de composición de experimentos que acabamos de estudiar resultan especialmente útiles para asignar probabilidades a los sucesos elementales, de forma que sea aplicable la Regla de Laplace, y podamos calcular la probabilidad de un suceso, dividiendo el número de casos favorables al mismo, por el de casos posibles.

En el capítulo anterior, señalamos que uno de los problemas que se presentan cuando tratamos de utilizar dicha regla es cómo escribir el espacio muestral de modo que pueda aceptarse que los sucesos elementales (los casos posibles) sean equiprobables. Veamos, con dos ejemplos, que las técnicas anteriores son una herramienta imprescindible para resolver dicho problema.

Ejemplo 3.12. Se lanza una moneda perfecta dos veces. Deseamos calcular la probabilidad de *obtener una o dos caras.*

Podemos considerar, al menos, dos espacios muestrales distintos para describir los resultados del experimento aleatorio:

$$\Omega_1 = \{(c,c),(c,z),(z,c),(z,z)\}$$

y

$$\Omega_2 = \{2 \text{ caras}, 1 \text{ cara}, 0 \text{ caras}\} .$$

En el primero, los resultados posibles son las cuatro variaciones con repetición de dos elementos (c y z, es decir, cara y cruz) tomados de dos en dos. En otras palabras, consideramos que son resultados distintos aquellos que proporcionan el mismo número de caras (y, por tanto, de cruces) en orden diferente. En el segundo espacio muestral, los resultados posibles son las tres combinaciones con repetición de dos elementos tomados de dos en dos. Este espacio sólo será utilizable si los sucesos por los que estamos interesados, como el planteado al comienzo del ejemplo, se refieren al número de caras y no al orden en que se obtienen (por ejemplo, no podríamos estudiar con Ω_2 el suceso *obtener cara en el primer lanzamiento*).

El segundo espacio muestral es, obviamente, más sencillo que el primero. Si, por ejemplo, la moneda se lanzase diez veces, Ω_1 tendría $2^{10} = 1024$ elementos, mientras que Ω_2 tendría únicamente once (los números naturales desde el cero hasta el diez).

Todo parece apuntar, en este caso, hacia la conveniencia de utilizar Ω_2. Ahora bien, no sólo tratamos de buscar un espacio muestral sencillo y que contenga al suceso de interés; buscamos, también, un espacio en el que se pueda aplicar la regla de Laplace, lo cual facilitará el cálculo de las probabilidades. En este sentido, ¿es razonable suponer que Ω_2 es equiprobable? Veamos, en primer lugar, que ambos espacios no pueden serlo simultáneamente. Así, si suponemos que los sucesos elementales de Ω_1 son equiprobables y, por tanto, su probabilidad es $1/4$, no lo serán los de Ω_2. Evidentemente, si queremos asegurar una coherencia en la asignación de probabilidades en Ω_1 y Ω_2, la probabilidad de obtener *0 caras* o la de obtener *2 caras* deberán ser iguales a $1/4$ (coincidirían, respectivamente, con las probabilidades de los sucesos $\{(z,z)\}$ y $\{(c,c)\}$ de Ω_1), mientras que la de obtener *1 cara* sería igual a $1/2$, ya que

$$p(1 \text{ cara}) = p[(c,z),(z,c)] = \frac{1}{2} .$$

Lo que ocurre no es sino un resultado conocido en el Análisis Combinatorio. El número de variaciones con repetición que están incluidas en una combinación con repetición es el

número de permutaciones con repetición que se puedan hacer a partir de la combinación; pero este número es diferente según cuántos elementos se repitan en la combinación. Por tanto, si las variaciones con repetición son equiprobables, cada combinación con repetición deberá tener una probabilidad proporcional al número de permutaciones con repetición que puedan generarse a partir de ella.

El comentario anterior sólo indica que los dos espacios muestrales no pueden ser simultáneamente equiprobables, pero no aclara cuál de ellos podría ser así considerado. Ahora bien, si descomponemos el experimento original en dos, consistentes cada uno de ellos en un lanzamiento de la moneda, y suponemos, como parece razonable, que dichos experimentos son independientes, las probabilidades de los sucesos elementales de Ω_1 serán:

$$\begin{aligned} p[(c,c)] = p_1(c)\cdot p_2(c) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}\,, \\ p[(c,z)] = p_1(c)\cdot p_2(z) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}\,, \\ p[(z,c)] = p_1(z)\cdot p_2(c) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4} \quad \text{y} \\ p[(z,z)] = p_1(z)\cdot p_2(z) &= \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}\,. \end{aligned}$$

En definitiva, los sucesos elementales de Ω_1 son equiprobables (no así los de Ω_2), y es en este espacio muestral donde la regla de Laplace es aplicable. De esta forma, y dado que el suceso *obtener al menos una cara* ocurre en tres de los cuatro casos posibles, su probabilidad será:

$$p(\text{obtener al menos una cara}) = \frac{3}{4}\,.$$

Digamos, finalmente, que, como el suceso en el que estamos interesados no alude al orden en que se obtienen las caras, podríamos, ahora, probabilizar los resultados elementales de Ω_2 a partir de los de Ω_1, y calcular en este espacio su probabilidad. ◁

Ejemplo 3.13. Se entregan cuatro cartas de una baraja española a un jugador. Deseamos calcular la probabilidad de que *todas ellas sean del mismo palo.*

Como en el ejemplo anterior, podemos construir dos espacios muestrales asociados a este experimento. El primero, Ω_1, tendría como resultados posibles las 2193360 variaciones sin repetición de 40 elementos (las cuarenta cartas) tomados de cuatro en cuatro, esto es, se considerarían resultados distintos los que proporcionasen las mismas cartas en distinto orden. El segundo espacio muestral, Ω_2, tendría como elementos las 91390 combinaciones sin repetición de las cuarenta cartas tomadas de cuatro en cuatro. En este caso, dos resultados distintos se diferenciarían, al menos, en una carta.

Como la probabilidad que deseamos calcular se refiere a las cartas entregadas, y no a su orden, podríamos utilizar, en principio, cualquiera de los dos espacios muestrales, si bien el hecho de no tener en cuenta el orden y su menor tamaño, parecen aconsejar la utilización de Ω_2. No obstante, la asignación de probabilidades se puede realizar con más facilidad en Ω_1. Si consideramos que entregar cuatro cartas a la vez equivale a entregar esas mismas cuatro cartas de una en una[2], el experimento puede descomponerse en cuatro experimentos más sencillos; cada uno de ellos consistiría en entregar una carta al jugador.

2. La práctica de considerar que una entrega simultánea de varios objetos equivale a la entrega sucesiva de cada uno de ellos o, dicho de otro modo, la de hacer equivalente la extracción simultánea de varios objetos de una urna a la extracción sucesiva sin reemplazamiento de cada uno de ellos, es habitual en el cálculo de probabilidades para ese tipo de experimentos, ya que dicho

Dichos experimentos no son independientes, aunque se conoce la relación que existe entre ellos.

Con las técnicas de composición de experimentos, podríamos calcular la probabilidad de un resultado cualquiera. Por ejemplo, la probabilidad de *obtener el as de oros, el tres de copas, el cinco de copas y el rey de bastos* (en este orden) sería

$$p_1(\text{as de oros}) \cdot p_2(\text{tres de copas}) \cdot p_3(\text{cinco de copas}) \cdot p_4(\text{rey de bastos}) = \\ = \frac{1}{40} \cdot \frac{1}{39} \cdot \frac{1}{38} \cdot \frac{1}{37} = \frac{1}{2193360} ,$$

donde p_i $(i = 1, \ldots, 4)$ es la probabilidad asignada al experimento consistente en extraer una carta de una baraja de la que se han retirado las $i - 1$ cartas previamente extraídas.

Obsérvese que la misma probabilidad se obtendría para cualquier otro elemento de Ω_1 y, en consecuencia, Ω_1 es equiprobable. Como es lógico, la probabilidad de uno cualquiera de los sucesos elementales de Ω_1 resulta ser la unidad dividida por el número de ellos.

Las probabilidades en Ω_2 se pueden calcular a partir de las asignadas a los elementos de Ω_1. Cada elemento de Ω_2 es una de las 91390 combinaciones de las cuarenta cartas tomadas de cuatro en cuatro. Como todas, y cada una de las combinaciones, dan lugar a veinticuatro variaciones (es decir, el número de permutaciones sin repetición de cuatro elementos), la probabilidad de una cualquiera de ellas será veinticuatro veces la probabilidad de una variación, esto es,

$$\frac{24}{2193360} = \frac{1}{91390} .$$

En definitiva, los elementos de Ω_2 son, también, equiprobables, dado que la probabilidad de uno cualquiera de ellos es la unidad dividida por su número; esto es,

$$\frac{1}{C_{40,4}} = \frac{1}{91390} .$$

Téngase en cuenta que hemos considerado cada suceso elemental de Ω_2 como un suceso compuesto de Ω_1, calculando su probabilidad como suma de probabilidades de sucesos elementales de Ω_1. En efecto,

$$\frac{P_4}{V_{40,4}} = \frac{1}{\frac{V_{40,4}}{P_4}} = \frac{1}{C_{40,4}} .$$

En este caso, los dos espacios son equiprobables porque los sucesos elementales de Ω_2 pueden considerarse como sucesos compuestos de Ω_1, con el mismo número de elementos cada uno de ellos. Finalmente, calculemos la probabilidad pedida, utilizando Ω_2. Como hay $C_{10,4} = 210$ posibilidades de que las cuatro cartas sean de un palo concreto, y existen cuatro palos en la baraja, la probabilidad de que *las cuatro cartas sean del mismo palo* valdrá:

$$p(\text{cuatro cartas del mismo palo}) = \frac{4 \cdot C_{10,4}}{C_{40,4}} = \frac{4 \cdot 210}{91390} = 0.00919 . \triangleleft$$

Por último, en los experimentos que se pueden descomponer en varias etapas sucesivas, el cálculo de probabilidades presenta una dificultad añadida: la visión

cálculo sólo se puede realizar razonablemente de esta forma, y proporciona, además, resultados que están de acuerdo con la intuición.

global de las distintas fases del experimento. Ayuda a conseguir esta visión la utilización de una técnica gráfica que se conoce como **diagrama de árbol**. El siguiente ejemplo describe esta técnica.

Ejemplo 3.14. Un hospital psiquiátrico cuenta con 100 pacientes repartidos en dos pabellones. El psiquiatra ha diagnosticado que sólo uno de ellos está loco, internándole en el pabellón I junto con 49 cuerdos. Una noche, 3 pacientes del pabellón I pasan al pabellón II. Apercibidos los loqueros, y antes de que el psiquiatra advierta la negligencia, devuelven a 3 internados elegidos aleatoriamente. Enterado el psiquiatra del desaguisado, ¿en qué pabellón deberá buscar al paciente loco?

Es evidente que no hay seguridad de que el loco se encuentre en uno u otro pabellón. A lo más, podremos calcular en cuál de ellos se encuentra con más probabilidad. Para facilitar este cálculo, recurramos a un diagrama de árbol (figura 3.4) que nos clarifique el trajín nocturno de locos y cuerdos.

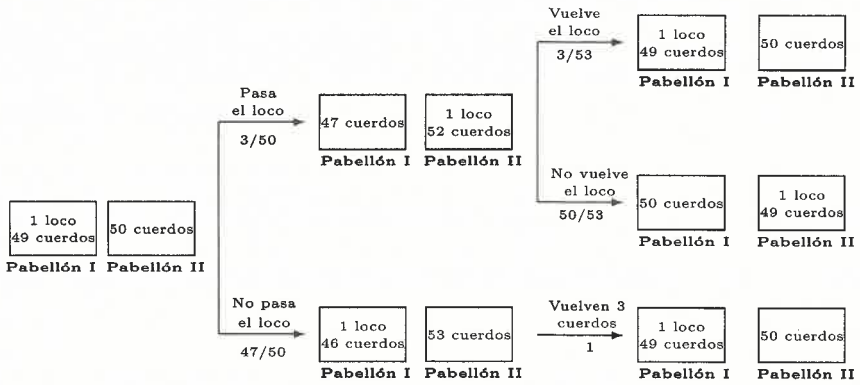


FIG. 3.4.

En primer lugar, calculemos la probabilidad de que el loco haya participado, o no, en la correría nocturna. Por comodidad, hallemos primero la probabilidad de que *no pase el loco*, es decir, la probabilidad de que *el loco no sea ni el primero, ni el segundo, ni el tercero de los pacientes tránsfugas.* Es decir:

$$p(\text{ no pase el loco }) = \frac{49}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} = \frac{47}{50}\,.$$

Evidentemente, la probabilidad de su complementario será:

$$p(\text{ pase el loco }) = \frac{3}{50}$$

Si el loco fue uno de los que pasó, es decir, si se encuentra en el pabellón II cuando los loqueros advierten la fuga, puede que sea uno de los tres que es devuelto (resolviéndose, aparentemente, el desaguisado), o puede, al no ser elegido, que permanezca en el pabellón que no le corresponde. Calculemos estas probabilidades, teniendo en cuenta que, en esta situación, hay 53 pacientes en el pabellón II, siendo uno de ellos el loco.

$$p(\text{ no devuelvan al loco }) = \frac{52}{53} \cdot \frac{51}{52} \cdot \frac{50}{51} = \frac{50}{53}$$

y

$$p(\text{ devuelvan al loco }) = 1 - \frac{50}{53} = \frac{3}{53} .$$

Nótese que éstas son realmente probabilidades de transición.

Si el loco no fue uno de los que pasó al pabellón II, es obvio que no podrá encontrarse en dicho pabellón.

Calculemos la probabilidad de que el loco esté en el pabellón I. Para ello, tengamos en cuenta las probabilidades de las *ramas* del *árbol* en cuyas *hojas* está el suceso en cuestión. Es decir:

$$p(\text{el loco está en el pabellón I}) = \frac{3}{50} \cdot \frac{3}{53} + \frac{47}{50} \cdot 1 = \frac{50}{53}$$

o bien,

$$p(\text{el loco está en el pabellón II}) = \frac{3}{50} \cdot \frac{50}{53} = \frac{3}{53} .$$

Señalemos que, en estos cálculos, no hemos aplicado otra cosa que el teorema de la probabilidad total. ◁

3.5 Ejercicios propuestos

3.1. Si los sucesos A y C cumplen que $A \cap B = C \cap B$, siendo B un suceso de probabilidad no nula, demuéstrese que

$$p(A \mid B) = p(C \mid B) .$$

3.2. Si A, B y C son tres sucesos tales que $p(B \mid A) \geq p(B)$ y $p(C \mid B) \geq p(C)$, pruébese con un ejemplo que la relación $p(C \mid A) \geq p(C)$ no tiene por qué ser cierta. (Sugerencia: considérese el experimento de lanzar un dado normal.)

3.3. Los sucesos A y B verifican que $p(A) = 1/4$, $p(B \mid A) = 1/2$ y $p(A \mid B) = 1/4$. Razónese si cada una de las siguientes afirmaciones es verdadera o falsa:

a) A y B son mutuamente excluyentes.
b) $p(\overline{A} \mid \overline{B}) = 3/4$.
c) $p(A \mid B) + p(A \mid \overline{B}) = 1$.

3.4. Sean A y B dos sucesos tales que $p(B) = 0.3$, $p(A \mid B) = 0.5$ y $p(A \cup B) = 0.6$. Calcúlese la probabilidad del suceso A.

3.5. Un individuo se dispone a realizar un viaje en avión desde Madrid a Damasco. Temeroso, se dirige a la compañía y pregunta cuál es la probabilidad de que haya una bomba en el avión. Siendo dicha probabilidad 0.1, y angustiado ante el riesgo, pregunta cuál sería la probabilidad de que haya dos bombas. Después de breves cálculos se le contesta que sería de 0.01. Más aliviado por la respuesta, decide llevar en su equipaje una bomba. ¿Qué opinión estadística merece dicha decisión?

3.6. Se sabe que en la mujeres embarazadas el resultado de la prueba de la rana es negativo el 5 % de los casos. Habiendo obtenido positivo, una mujer sospecha que está embarazada con probabilidad 0.95. ¿Qué juicio merece esta sospecha?

3.7. En una urna hay un total de 12 bolas entre blancas y negras. Sabiendo que la probabilidad de elegir dos bolas blancas en dos extracciones sin reemplazamiento es 1/11, ¿cuántas bolas negras hay en la urna?

3.8. En una estantería hay 50 libros de Historia y 62 de Estadística. Se eligen dos libros al azar, ¿cuál es la probabilidad de que ambos sean de Estadística? Si sabemos que uno es de Estadística, ¿cuál es la probabilidad de que ambos lo sean?

3.9. El 40 % de los alumnos de C.O.U. son de ciencias, el 25 % de letras y el resto recibe enseñanza mixta. El porcentaje de mujeres en cada uno de estos grupos es 40, 60 y 55, respectivamente. ¿Qué proporción de mujeres son del grupo de ciencias?

3.10. La constructora CEMENTASA trata de determinar si debería presentar una oferta para la construcción de una nueva autovía, pues su principal competidora, la constructora ARENAL, ha conseguido un 70 % de las adjudicaciones en el pasado. Si ARENAL no presenta una oferta, la probabilidad de que su rival obtenga la obra es 0.5; si presenta oferta, la probabilidad de que CEMENTASA reciba el contrato es 0.25.

a) Si la constructora CEMENTASA gana el contrato, ¿cuál es la probabilidad de que la constructora ARENAL no haya presentado oferta?

b) ¿Cuál es la probabilidad de que la constructora CEMENTASA gane el contrato?

3.11. Tres compañías de seguros copan el mercado de una determinada ciudad. El 30 % de las pólizas suscritas corresponden a la compañía A, el 25 % a la B y el 45 % restante a la C. El porcentaje de pólizas de seguros de vida en cada una de ellas es del 15, 20 y 25 % , respectivamente.

a) ¿Qué proporción de las pólizas suscritas corresponde a seguros de vida?

b) Un individuo ha suscrito un seguro de vida. ¿Cuál es la probabilidad de que su póliza sea de la primera entidad?

3.12. Juan dice la verdad nueve veces de cada diez, y Pedro siete veces de cada nueve. Se extrajo una bola al azar de una bolsa que contenía cinco bolas blancas y veinte negras. Ambos dijeron que la bola extraída era blanca. ¿Cuál es la probabilidad de que la bola extraída fuera realmente blanca?

3.13. Un auditor ha sido informado de la existencia de un error deliberado en uno de los documentos del departamento de contabilidad de una empresa.
La probabilidad de que el error se encuentre en un lote de 10 documentos es 0.05, siendo igualmente probable que el defecto esté en cualquiera de estos documentos. Si ha inspeccionado ya 9 sin detectar el fallo, ¿cuál es la probabilidad de que el fraude no se encuentre en dicho lote?

3.14. Sean A y B sucesos tales que $p(A) = 0.3$, $p(A \cup B) = 0.8$ y $p(B) = p$.

a) ¿Para qué valor de p son A y B independientes?

b) Obténgase una condición necesaria para que A y B sean disjuntos.

3.15. Sean A y B dos sucesos disjuntos y de probabilidad no nula. Demuéstrese que A y B son sucesos dependientes.

3.16. Sean A, B y C sucesos independientes. Pruébese:

a) A y $B \cup C$ son independientes.
b) A y $B \cap C$ son independientes.

3.17. Sean A y B dos sucesos independientes con probabilidad no nula. Compruébese que para cualquier suceso C se verifica la relación

$$p(C \mid A) = p(B) \cdot p(C \mid AB) + p(\overline{B}) \cdot p(C \mid A\overline{B}) .$$

3.18. Se lanzan tres monedas perfectas y se consideran los sucesos A, *obtener el mismo resultado con la primera y la segunda*, B, *obtener el mismo resultado con la segunda y la tercera*, y C, *obtener el mismo resultado con la primera y la tercera*. Demuéstrese que los sucesos A, B y C son independientes dos a dos, pero no mutuamente independientes.

3.19. Sean A, B, C y D cuatro sucesos mutuamente independientes. Demuéstrese que también lo son:

a) A y $B \cup C$.
b) $A \cap C$ y $B \cup \overline{D}$.
c) $A \cap B$, C y $\overline{D}$.

3.20. Sean dos sucesos A y B tal que $0 < p(B) < 1$. Demuéstrese que A y B son independientes si y sólo si

$$p(A \mid B) = p(A \mid \overline{B}) .$$

3.21. En el espacio muestral $\Omega = (0, 1)$, consideremos la familia de sucesos formada por los subintervalos de $(0, 1)$ de la forma (a, b), siendo a y b números reales tales que $0 < a < b < 1$.
Si la probabilidad de un suceso de este tipo viene dada por la amplitud del subintervalo, es decir, $p[(a, b)] = b - a$, hállense dos sucesos que sean independientes.

3.22. Sea $\Omega = (0, 2)$ un espacio muestral en el que se define la probabilidad de un subintervalo (a, b) de Ω como el área comprendida entre el propio subintervalo (a, b) y la recta $2y = x$.
Encuéntrense dos sucesos cuya probabilidad sea mayor que cero y menor que uno, que sean independientes.

3.23. Sean A y B dos sucesos cualesquiera de $(\Omega_1, \alpha_1, p_1)$ y $(\Omega_2, \alpha_2, p_2)$, espacios probabilísticos correspondientes a dos experimentos independientes. Sea (Ω, α, p) el espacio probabilístico asociado al experimento compuesto. Demuéstrese que

$$p(A \times B) = p_1(A) \cdot p_2(B) .$$

3.24. Pruébese que la probabilidad dada en la definición 3.8 es la única que verifica las proposiciones 3.8 y 3.9.

3.25. Demuéstrese que la probabilidad dada en la definición 3.9 es la única que verifica la proposición 3.10.

3.26. Una moneda está trucada de modo que la probabilidad de cara es el doble que la de cruz. Hállese la probabilidad de no obtener ninguna cara en una serie de 10 tiradas con dicha moneda.

3.27. Una empresa tiene preferencia por los varones a la hora de elegir sus empleados; en concreto, en el 70 % de las contrataciones se ha elegido un varón. Se ha formado un comité de seguridad e higiene formado por 20 personas elegidas aleatoriamente. ¿Cuál es la probabilidad de que al menos la mitad sean hombres?

3.28. Un jugador, de los cinco que van a disputar una partida con una baraja española, reparte todas las cartas. ¿Cuál es la probabilidad de que reciba al menos seis cartas del mismo palo?

3.29. En un cajón hay 10 bombillas, 4 de ellas fundidas. Una lámpara de dos focos alumbra sólo cuando ambas bombillas están en buen estado. ¿Cuál es la probabilidad de que la lámpara no funcione tras haberle colocado dos bombillas elegidas al azar del cajón?

3.30. En una pecera hay 10 peces rojos y 8 verdes. Un niño travieso captura con una red 5 de ellos. ¿Cuál es la probabilidad de que haya 4 del mismo color?

3.31. Una empresa cuenta con tres máquinas. La máquina A no tiene ningún sistema de seguridad, dejando de funcionar cuando se presenta el primer fallo en el mecanismo. En cambio, las máquinas B y C tienen un sistema de seguridad que les permite resistir uno o dos fallos, respectivamente, antes de dejar de funcionar. Se supone que la probabilidad de que falle el mecanismo es la misma en las tres máquinas. Sabiendo que hasta el momento sólo se ha detectado un fallo en la máquina B, ¿cuál es la probabilidad de que la primera máquina que deje de funcionar sea la C?

3.32. Demuéstrese que es más probable obtener por lo menos un as con cuatro dados, que por lo menos un doble as en veinticuatro tiradas de doce dados. (Paradoja de Méré.)

3.33. Una urna contiene seis bolas blancas, tres negras y dos rojas. Dos de las bolas blancas están numeradas. Se sacan dos bolas de la urna:

a) ¿Cuál es la probabilidad de que las dos sean del mismo color?
b) Si son del mismo color, ¿cuál es la probabilidad de que sean blancas?
c) Si son blancas, ¿cuál es la probabilidad de que al menos una esté numerada?

3.34. Se tira un dado normal dos veces, ¿cuál es la probabilidad de que la suma de las puntuaciones sea un número par?; ¿y la probabilidad de que al menos una de las puntuaciones sea mayor que cuatro?

3.35. Cada vez que un cliente compra un dentífrico elige la marca A o la marca B. Supóngase que la probabilidad de que elija la misma marca que en la compra

anterior es $1/3$. Si la primera marca que adquiere es A, ¿cuál es la probabilidad de que la quinta compra sea de la marca B?

3.36. Tres aventureros deben elegir a uno de ellos para una misión arriesgada. Para ello, toman una urna con dos bolas blancas y una negra y extraen cada uno sucesivamente una bola sin reemplazamiento. El elegido será el que escoja la bola negra. Demuéstrese que todos tienen la misma probabilidad de que les toque.

3.37. Un contable tiene sobre su mesa dos grupos de 20 facturas cada uno. En el primer lote hay dos facturas con errores de cálculo y en el segundo tres.
Una corriente de aire hace que las facturas caigan de la mesa y, al recogerlas, una del primer grupo se confunde entre las del segundo. ¿Cuál es la probabilidad de que, al revisar una factura del segundo grupo, tenga un error?

3.38. Un test rápido sobre el contenido alcohólico en la sangre de un conductor, realizado por la policía en la carretera, es fiable en el 80 % de las ocasiones. Los conductores que dan positivo son sometidos por un médico a un test más preciso, que nunca falla en un conductor sobrio, pero que tiene un 10 % de error en los embriagados. Sabiendo que el 5 % de los conductores detenidos por la policía está embriagado:

a) ¿Qué proporción de conductores detenidos será sometida a un segundo test?

b) Un conductor dio negativo en el segundo test. ¿Cuál es la probabilidad de que condujese con un contenido de alcohol en sangre mayor del legal?

Capítulo 4

VARIABLES ALEATORIAS UNIDIMENSIONALES

Hasta ahora, nos hemos dedicado a estudiar las propiedades generales más importantes de los espacios probabilísticos. Sin embargo, con frecuencia, no interesa estudiar los propios sucesos de una σ-álgebra, sino una o varias características numéricas ligadas a los resultados posibles del experimento.

Cuando realizamos el experimento consistente en lanzar una moneda dos veces y observar el resultado obtenido, no estamos tan interesados en la probabilidad de un suceso cualquiera (por ejemplo, obtener primero una cara y luego una cruz), como en la de conseguir un número determinado de caras. Asimismo, cuando elegimos un individuo al azar de una población, nos puede interesar la probabilidad de que su renta anual no sea superior a tres millones de pesetas; o, seleccionada una empresa entre las que forman el censo industrial de un país, nuestro interés puede estar en la probabilidad de que el número de sus trabajadores sea mayor que diez y menor que cincuenta.

Estas características numéricas (número de caras, renta anual, número de trabajadores, etc.) a las que, bajo ciertas condiciones, podemos asociar una probabilidad inducida por la probabilidad de los sucesos del experimento, reciben el nombre de variables aleatorias y serán objeto de estudio a lo largo de los siguientes capítulos.

De esta forma, las probabilidades, que hasta ahora definíamos para sucesos (subconjuntos de un cierto espacio Ω), se transformarán en probabilidades definidas en subconjuntos de la recta real. Más aún, en última instancia, podremos caracterizar las probabilidades inducidas por las variables aleatorias mediante funciones reales de variable real, con las ventajas que esto supone desde el punto de vista del tratamiento matemático.

4.1 Definición de variable aleatoria

La importancia y complejidad de las definiciones que enunciaremos a lo largo de los siguientes apartados, hacen conveniente el desarrollo previo de un ejemplo que permita intuir los principales conceptos y su necesidad.

Ejemplo 4.1. Realicemos el experimento consistente en lanzar una moneda equilibrada dos veces y observar el resultado. El espacio muestral asociado será:

$$\Omega = \{(c,c),(c,z),(z,c),(z,z)\}\ ,$$

donde designamos por c el resultado *obtener cara* y por z *obtener cruz.* Teniendo en cuenta la naturaleza de la moneda, podemos suponer que todos los resultados son igualmente probables, es decir,

$$p[(c,c)] = p[(c,z)] = p[(z,c)] = p[(z,z)] = 1/4\ ,$$

lo que implica que estamos trabajando con la σ-álgebra $\alpha = \mathcal{P}(\Omega)$.

Ahora bien, nos puede interesar estudiar *el número de caras obtenidas*, característica numérica de los resultados del experimento que designaremos por X. Podemos considerar X como una aplicación que a cada resultado del experimento le hace corresponder un número real, el número de caras que se observan.

$$\begin{array}{rccc} X: & \Omega & \to & \mathbb{R} \\ & (c,c) & \hookrightarrow & 2 \\ & (c,z) & \hookrightarrow & 1 \\ & (z,c) & \hookrightarrow & 1 \\ & (z,z) & \hookrightarrow & 0 \end{array}$$

Entendiendo la probabilidad de obtener una cara como la probabilidad del conjunto de resultados del experimento en los que se observa una sola cara, podemos escribir:

$$p[X=1] = p[(c,z),(z,c)] = p[(c,z)] + p[(z,c)] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}\ .$$

De igual forma, obtendríamos otras probabilidades inducidas por la variable:

$$p[X \geq 1] = p[(c,z),(z,c),(c,c)] = 3/4$$

y

$$p[X < 2] = p[(z,z),(c,z),(z,c)] = 3/4\ .$$

De este modo, hemos trasladado la probabilidad definida en los sucesos del experimento a subconjuntos de la recta real del tipo $\{1\},[1,\infty),(-\infty,2),\ldots$ $\triangleleft$

Por tanto, trataremos de probabilizar una característica numérica referente a los resultados de un cierto experimento, induciendo una probabilidad sobre la recta real $\mathbb{R}$.

Definición 4.1 *Sea (Ω,α,p) un espacio probabilístico. Una aplicación*

$$X : \Omega \longrightarrow I\!R$$

se denomina variable aleatoria si

$$X^{-1}((-\infty,x]) \in \alpha \text{ para todo número real } x.$$

No es imprescindible definir una variable aleatoria sobre un espacio probabilístico, (Ω,α,p). Sería suficiente con que dicho espacio fuera probabilizable. Ahora bien, dado que más adelante asignaremos probabilidades a los valores de la variable

a partir de las probabilidades definidas en el espacio muestral Ω, es conveniente considerar definida, de antemano, la variable sobre un espacio probabilístico.

OBSERVACIONES

1) Nótese que las variables aleatorias no son sino aplicaciones, no todas, del espacio muestral en $\mathbb{R}$. Son, por tanto, asignaciones de valores que se realizan en un espacio muestral Ω. En otras palabras, envían cada elemento ω del espacio muestral a un número real que escribiremos, con la terminología usual de las aplicaciones, como $X(\omega)$.

Por motivos históricos, se utiliza el nombre de *variables aleatorias* en vez del de *funciones*, que es lo que son realmente; consecuentemente, en lugar de representarse con las letras minúsculas f, g o h, como es usual, se describen mediante las últimas letras del abecedario, U, V, W, X, Y o Z, siempre en mayúsculas.

El adjetivo *aleatorias* que acompaña a su denominación sirve para diferenciarlas de su antecedente histórico, las variables estadísticas, que son el objeto de estudio de la Estadística Descriptiva. Las variables estadísticas recogían una característica numérica observada de los individuos de una población; en este sentido, se contabilizaban los valores observados en términos de frecuencia relativa. Como veremos más adelante, los posibles valores de una variable aleatoria vendrán contabilizados en términos de probabilidad.

2) Evidentemente, no todas las aplicaciones de Ω en $\mathbb{R}$ son variables aleatorias. La definición exige que $X^{-1}((-\infty, x])$ pertenezca a la σ-álgebra α, es decir, que $X^{-1}((-\infty, x])$ sea un suceso del espacio probabilístico (Ω, α, p), y ello para cualquier número real x. El conjunto $X^{-1}((-\infty, x])$ es la contraimagen, o imagen recíproca, del intervalo de números reales $(-\infty, x]$.

Recordemos que, si f es una aplicación de un conjunto M en otro, N,

$$f : M \to N \, ,$$

entonces f^{-1}, que se denomina imagen recíproca, es una aplicación de $\mathcal{P}(N)$ en $\mathcal{P}(M)$,

$$f^{-1} : \mathcal{P}(N) \to \mathcal{P}(M) \, ,$$

tal que

$$f^{-1}(B) = \{a \in M \mid f(a) \in B\}$$

para todo B subconjunto de N.

En definitiva, f^{-1} (que no es, evidentemente, $\frac{1}{f}$) asocia a cada subconjunto B de N el subconjunto de M formado por los puntos que f aplica en B, como se representa esquemáticamente en la figura 4.1.

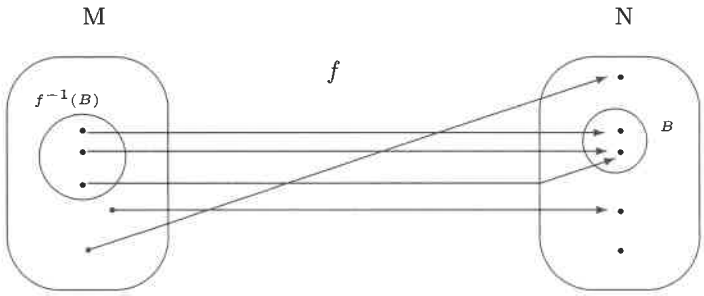


FIG. 4.1.

Por tanto, en nuestro caso,

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \in (-\infty, x]\} = \{\omega \in \Omega \mid X(\omega) \leq x\} ,$$

es decir, es el subconjunto de Ω formado por aquellos resultados del experimento, ω, tales que $X(\omega) \leq x$.

Para simplificar las expresiones que, como puede observarse son muy farragosas, fijaremos desde ahora la siguiente notación:

Sea X una variable aleatoria y B un subconjunto de $\mathbb{R}$. Con la notación

$$[X \in B]$$

queremos indicar el suceso de Ω

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} .$$

En particular, el suceso $X^{-1}((-\infty, x])$ lo escribiremos como $[X \leq x]$.

La figura 4.2 representa gráficamente la contraimagen del conjunto $(-\infty, 1.5]$ para la variable X del ejemplo 4.1. En concreto,

$$X^{-1}((-\infty, 1.5]) = [X \leq 1.5] = \{(c, z), (z, c), (z, z)\} .$$

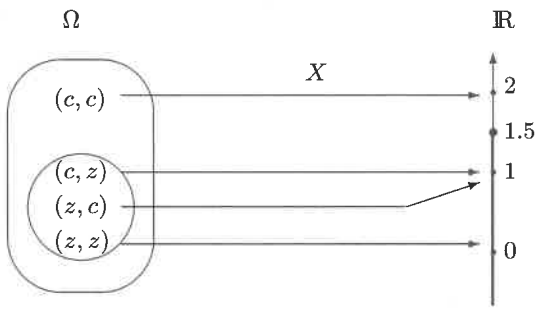


FIG. 4.2.

Se comprueba que X^{-1} es una aplicación que conserva las uniones, intersecciones y complementarios, es decir, que cumple

$$\begin{aligned} X^{-1}\left(\bigcup B_i\right) &= \bigcup X^{-1}(B_i)\,, \\ X^{-1}\left(\bigcap B_i\right) &= \bigcap X^{-1}(B_i) \text{ y} \\ X^{-1}(\overline{B}) &= \overline{X^{-1}(B)}\,, \end{aligned}$$

donde los conjuntos de $\{B_i\}$ y B son subconjuntos de $\mathbb{R}$. De hecho, esta propiedad no es exclusiva de las variables aleatorias, sino que la cumple cualquier aplicación.

Lema 4.1 *Sea una aplicación f de M en N. Entonces*

1. $f^{-1}\left(\bigcup B_i\right) = \bigcup f^{-1}(B_i)$.
2. $f^{-1}\left(\bigcap B_i\right) = \bigcap f^{-1}(B_i)$.
3. $f^{-1}(\overline{B}) = \overline{f^{-1}(B)}$.

donde los conjuntos de $\{B_i\}$ y B son subconjuntos cualesquiera de N.

Demostración. Sea $x \in f^{-1}\left(\bigcup B_i\right)$. Entonces $f(x) \in \bigcup B_i$ y, por tanto, $f(x)$ pertenece a alguno de los conjuntos de $\{B_i\}$, por ejemplo, a B_{i_0}. En consecuencia, $x \in f^{-1}(B_{i_0})$, con lo cual $x \in \bigcup f^{-1}(B_i)$. Así hemos demostrado que $f^{-1}\left(\bigcup B_i\right) \subset \bigcup f^{-1}(B_i)$.

Con un razonamiento análogo se demuestra que, si $x \in \bigcup f^{-1}(B_i)$, entonces $x \in f^{-1}(\bigcup B_i)$, es decir, $\bigcup f^{-1}(B_i) \subset f^{-1}\left(\bigcup B_i\right)$. Por tanto, ambos subconjuntos de M son iguales.

Probemos el segundo resultado (¿se atrevería el lector a probarlo una vez vista la demostración del primero?). Sea $x \in f^{-1}\left(\bigcap B_i\right)$. Entonces $f(x) \in \bigcap B_i$ y, por tanto, $f(x)$ pertenece a todos los conjuntos de $\{B_i\}$. En consecuencia, $x \in f^{-1}(B_i)$ para todo B_i, de lo que se deduce que $x \in \bigcap f^{-1}(B_i)$. En definitiva, $f^{-1}\left(\bigcap B_i\right) \subset \bigcap f^{-1}(B_i)$.

De forma análoga se demuestra que, si $x \in \bigcap f^{-1}(B_i)$, entonces $x \in f^{-1}\left(\bigcap B_i\right)$, es decir, $\bigcap f^{-1}(B_i) \subset f^{-1}\left(\bigcap B_i\right)$. Por tanto, los dos subconjuntos de M coinciden.

Para probar el tercer resultado partimos de que $x \in f^{-1}(\overline{B})$. Entonces, $f(x) \in \overline{B}$, o lo que es lo mismo, $f(x) \notin B$. Por tanto, $x \notin f^{-1}(B)$ y, consecuentemente, pertenece a $\overline{f^{-1}(B)}$. Es decir, $f^{-1}(\overline{B}) \subset \overline{f^{-1}(B)}$.

La contención recíproca se demuestra utilizando los mismos argumentos. ♣

3) ¿Por qué exigimos la condición de que $X^{-1}((-\infty, x])$ pertenezca a α para todo número real x, a las variables aleatorias? Sencillamente, para que dichos conjuntos sean sucesos y podamos, por tanto, probabilizarlos. Lo que a nosotros nos interesa desde un punto de vista probabilístico no son los sucesos de Ω, sino una característica numérica de los mismos; por ejemplo, la renta de una familia elegida al azar en una población. Nos puede interesar calcular la probabilidad de que esa renta no rebase el millón, o de que se encuentre entre dos y tres millones. Para calcular la primera de esas probabilidades, buscamos a las familias cuya renta no supere el millón y calculamos la probabilidad de que sea una de ellas la que

se obtiene al azar. En un lenguaje más formalizado, si Ω está formado por todas las familias de la población y X es la variable aleatoria *renta anual familiar en millones de pesetas*, es decir, la aplicación que a cada familia, $\omega \in \Omega$, le asocia el número real $X(\omega)$, su renta, se trataría de obtener $X^{-1}((-\infty, 1])$; esto es, el subconjunto de la población con renta menor o igual que un millón. Definiríamos, entonces, la probabilidad de que la renta sea menor o igual que un millón como

$$p[X^{-1}((-\infty, 1])] ,$$

es decir, la probabilidad del subconjunto de Ω formado por las familias con renta menor o igual que un millón, que, abreviadamente, escribiremos como $p[X \leq 1]$. Obviamente, esta forma razonable de trabajar exige que la probabilidad anterior sea calculable, o dicho de otra forma, que $X^{-1}((-\infty, 1])$ sea un suceso para el que hayamos definido una probabilidad, por tanto, un suceso de α.

En este sentido, podemos entender una variable aleatoria como la característica numérica de un individuo que se puede probabilizar.

Ejemplo 4.2. Una urna contiene tres bolas de distintos colores: blanco, rojo y negro. Consideremos la variable aleatoria X, *número total de bolas blancas* obtenidas en dos extracciones con reemplazamiento. El espacio muestral asociado al experimento es:

$$\Omega = \{(b,b),(b,r),(b,n),(r,b),(r,r),(r,n),(n,b),(n,r),(n,n)\} ,$$

donde b, r y n indican, respectivamente, sacar bola blanca, roja y negra. Parece razonable asignar a cada uno de los sucesos elementales una misma probabilidad, igual a $1/9$. La variable aleatoria X es una aplicación del espacio muestral Ω en $\mathbb{R}$ tal que

$$X[(b,b)] = 2 ,$$

$$X[(b,r)] = X[(b,n)] = X[(r,b)] = X[(n,b)] = 1 \quad \text{y}$$
$$X[(r,r)] = X[(r,n)] = X[(n,r)] = X[(n,n)] = 0 .$$

Por tanto, el conjunto de valores de la variable X es:

$$X(\Omega) = \{0,1,2\} .$$

Como hemos probabilizado todos los sucesos elementales y Ω es finito, la σ-álgebra sobre Ω es $\mathcal{P}(\Omega)$. Comprobaremos, utilizando la definición, que X es realmente una variable aleatoria. Para ello, hallaremos las contraimágenes por X de los intervalos $(-\infty, x]$. Es decir,

$$X^{-1}((-\infty,x]) = \begin{cases} \emptyset & x < 0 \\ \{(r,r),(r,n),(n,r),(n,n)\} & 0 \leq x < 1 \\ \{(b,r),(b,n),(r,b),(r,r),(r,n),(n,b),(n,r),(n,n)\} & 1 \leq x < 2 \\ \Omega & x \geq 2 \end{cases}$$

Evidentemente, $[X \leq x]$ pertenece a la σ-álgebra $\mathcal{P}(\Omega)$, sea cual sea x. Por tanto, X será una variable aleatoria y podremos calcular la probabilidad de cualquier conjunto $[X \leq x]$. Por ejemplo,

$$p[X \leq 1] = p[(b,r),(b,n),(r,b),(r,r),(r,n),(n,b),(n,r),(n,n)] = 8/9 .$$

La variable aleatoria X, *número total de bolas blancas*, la hemos considerado en el espacio probabilístico (Ω, α, p) donde $\alpha = \mathcal{P}(\Omega)$. Supongamos que construimos esta variable en un espacio probabilístico con otra σ-álgebra. En concreto, supongamos que sólo estamos interesados en que esta σ-álgebra, α', contenga al suceso B, *obtener al menos una bola blanca*, es decir:

$$B = \{(b,b),(b,r),(b,n),(r,b),(n,b)\} .$$

En ese caso, la σ-álgebra será:

$$\alpha' = \{\emptyset, B, \overline{B}, \Omega\} .$$

Si mantenemos la idea de equiprobabilidad, pero no su formalización (los sucesos elementales no son sucesos de α' y, por tanto, no tienen asignada probabilidad), podemos dar la siguiente probabilidad en α':

$$\begin{aligned} p'(\emptyset) &= 0 , \\ p'(B) &= 5/9 , \\ p'(\overline{B}) &= 4/9 \text{ y} \\ p'(\Omega) &= 1 . \end{aligned}$$

Veamos que X no es ahora una variable aleatoria. Evidentemente

$$X^{-1}((-\infty, 1]) = [X \leq 1] \notin \alpha' .$$

En este espacio probabilístico, (Ω, α', p'), no podemos probabilizar la característica numérica X, es decir, ésta no es una variable aleatoria. ◁

4) El problema que la condición de la definición de variable aleatoria trata de resolver, y que hemos analizado en la observación anterior, podría solucionarse de otra forma. En efecto, si lo que nos interesa probabilizar no es Ω, sino una característica ligada a los resultados de Ω, ¿por qué no definir directamente como resultados posibles del experimento los valores que toma esa característica? Es decir, ¿por qué no tomar

$$\Omega' = \{X(\omega) \text{ para todo } \omega \in \Omega\}$$

como espacio muestral?

Aunque este modo de trabajar no es, en absoluto, incorrecto, sin embargo suele ser más fácil probabilizar los sucesos de Ω que los de Ω'; además, probabilizado Ω, y actuando como lo hicimos en la observación anterior, podremos probabilizar de forma sencilla sucesos relacionados con varias características, sin tener que plantearse cada vez un nuevo espacio probabilístico.

Ejemplo 4.3. Se realizan dos tiradas de un dado perfecto. Estamos interesados en estudiar la suma de las puntuaciones.

El espacio muestral es

$$\Omega = \{(i,j); i = 1,\ldots,6, j = 1,\ldots,6\} ,$$

y cada uno de los sucesos elementales tiene una probabilidad igual a $1/36$.

Sea X la variable aleatoria *suma*, es decir, la aplicación de Ω en $\mathbb{R}$ tal que

$$X[(i,j)] = i + j, \quad i = 1, \ldots, 6, \; j = 1, \ldots, 6 \, .$$

Como hemos probabilizado todos los sucesos elementales y Ω es finito, la σ-álgebra sobre Ω es $\mathcal{P}(\Omega)$. El lector puede comprobar que X cumple la condición exigida en la definición, por lo que será una variable aleatoria.

Supongamos que queremos calcular, por ejemplo, la probabilidad de que la suma sea menor o igual que cinco; dicho suceso puede escribirse como

$$\begin{aligned} [X \leq 5] &= \{(i,j) \in \Omega \mid X[(i,j)] \leq 5\} = \\ &= \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\} \end{aligned}$$

y, por tanto, la probabilidad buscada valdrá

$$p[X \leq 5] = 10 \cdot \frac{1}{36} = 0.2778 \, .$$

Otra manera de proceder sería la de tomar las sumas como resultados posibles del experimento, es decir, considerar como espacio muestral

$$\Omega' = \{2,3,4,5,6,7,8,9,10,11,12\} \, .$$

Ahora bien, el cálculo de las probabilidades de los once sucesos elementales de Ω' no se habría simplificado (de hecho, deberíamos recurrir, explícita o implícitamente, a las probabilidades definidas sobre Ω). Además, los resultados obtenidos no nos servirían para estudiar otras características (salvo, obviamente, para las que se obtengan a partir de la suma). $\triangleleft$

La posibilidad de que una aplicación X de Ω en $\mathbb{R}$ sea una variable aleatoria, depende del *tamaño* de la σ-álgebra α. En efecto, cuantos más sucesos haya en la σ-álgebra, más fácil es que se cumpla la condición que definía una variable aleatoria. Esta idea resultará más clara a la vista de los siguientes resultados.

Proposición 4.1 *Sea (Ω, α, p) un espacio probabilístico y X una aplicación de Ω en $\mathbb{R}$. Entonces:*

1. *Si $\alpha = \mathcal{P}(\Omega)$, X es siempre una variable aleatoria.*
2. *Si X es una aplicación constante, es una variable aleatoria sea cual sea la σ-álgebra α.*
3. *Si $\alpha = \{\emptyset, \Omega\}$, X es una variable aleatoria si y sólo si es una aplicación constante.*

Demostración. Probaremos sólo la primera afirmación, dejando la demostración de las otras dos como un ejercicio para el lector.

La contraimagen $X^{-1}((-\infty, x])$ está contenida en Ω y, por tanto, pertenece a $\alpha = \mathcal{P}(\Omega)$ para todo número real x; en consecuencia, X es una variable aleatoria. $\clubsuit$

Veremos, a continuación, un resultado que nos ayudará a manejar las variables aleatorias.

Proposición 4.2 *Sea (Ω, α, p) un espacio probabilístico y X una aplicación de Ω en $\mathbb{R}$. Las siguientes condiciones son equivalentes:*

1. $X^{-1}((-\infty, x]) \in \alpha$ *para todo* $x \in \mathbb{R}$.
2. $X^{-1}((-\infty, x)) \in \alpha$ *para todo* $x \in \mathbb{R}$.
3. $X^{-1}((x, \infty)) \in \alpha$ *para todo* $x \in \mathbb{R}$.
4. $X^{-1}([x, \infty)) \in \alpha$ *para todo* $x \in \mathbb{R}$.

Demostración. Puesto que $X^{-1}((-\infty, x])$ y $X^{-1}((x, \infty))$ son complementarios, del mismo modo que $X^{-1}((-\infty, x))$ y $X^{-1}([x, \infty))$, las condiciones *1* y *3* son equivalentes entre sí, así como las condiciones *2* y *4*.

Veamos que las condiciones *1* y *2* son equivalentes, siéndolo, en consecuencia, las cuatro condiciones enunciadas. En efecto, podemos escribir

$$(-\infty, x] = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right)$$

y, por tanto, teniendo en cuenta el Lema 4.1,

$$X^{-1}((-\infty, x]) = \bigcap_{n=1}^{\infty} X^{-1}\left(\left(-\infty, x + \frac{1}{n}\right)\right) .$$

Así, deducimos que $X^{-1}((-\infty, x])$ se puede escribir como intersección de una colección infinita numerable de sucesos de α (por la condición *2*), con lo cual pertenece a α. Por tanto, queda probado que la condición *2* implica la *1*.

De forma análoga se comprueba que la condición *1* implica la *2*; en efecto,

$$(-\infty, x) = \bigcup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n}\right] .$$

A partir de aquí, la demostración sigue los razonamientos de la anterior.

En definitiva, las condiciones *1* y *2* son equivalentes y, por consiguiente, lo son, también, las cuatro enunciadas. ♣

La proposición anterior tiene consecuencias importantes. En efecto, dado un espacio probabilístico (Ω, α, p) y una variable aleatoria X, vamos a poder probabilizar los intervalos de números reales de la forma $(-\infty, x]$, $(-\infty, x)$, (x, ∞) y $[x, \infty)$, o, con mayor precisión, la contraimagen por X de éstos. Es decir, podremos calcular las siguientes probabilidades:

$$p[X^{-1}(-\infty, x]], \quad p[X^{-1}(-\infty, x)], \quad p[X^{-1}(x, \infty)] \quad \text{y} \quad p[X^{-1}[x, \infty)] ,$$

o, lo que es lo mismo,

$$p[X \leq x], \quad p[X < x], \quad p[X > x] \quad \text{y} \quad p[X \geq x] .$$

Pero no son éstos los únicos intervalos que podemos probabilizar. De hecho, la proposición anterior nos va a permitir probabilizar cualquier intervalo real (más en

concreto, la contraimagen por X de cualquier intervalo). Por ejemplo, si a y b son dos números reales con $a < b$, podemos calcular la probabilidad de que la variable aleatoria *se encuentre* entre a y b, ya que $X^{-1}([a,b])$ es un suceso. En efecto, como

$$(-\infty, b] = (-\infty, a) \cup [a, b],$$

podemos escribir,

$$X^{-1}((-\infty, b]) = X^{-1}((-\infty, a)) \cup X^{-1}([a, b]) .$$

Por tanto, $X^{-1}([a,b]) \in \alpha$ ya que

$$X^{-1}([a,b]) = X^{-1}((-\infty, b]) - X^{-1}((-\infty, a)) ,$$

es decir, se puede poner como diferencia de sucesos de α. Además,

$$p[X^{-1}((-\infty, b])] = p[X^{-1}((-\infty, a))] + p[X^{-1}([a, b])]$$

con lo que

$$p[X^{-1}([a,b])] = p[X^{-1}((-\infty, b])] - p[X^{-1}((-\infty, a))] .$$

Expresándolo en la notación abreviada,

$$p[a \leq X \leq b] = p[X \leq b] - p[X < a] .$$

Análogamente, es posible calcular $p[a < X \leq b], p[a < X < b]$ o $p[a \leq X < b]$.

De hecho, la proposición 4.2 se puede ampliar introduciendo cuatro nuevas condiciones, equivalentes entre sí, y equivalentes a aquéllas: la contraimagen por X de cualquier intervalo finito de números reales ($[a,b]$, $(a,b]$, (a,b) y $[a,b)$) pertenece a la σ-álgebra α. Todo ello nos permitirá definir *una variable aleatoria como una aplicación de Ω en $\mathbb{R}$ que verifica que $X^{-1}(I) \in \alpha$, para todo intervalo real I.* Aun así, según veremos más adelante, será la condición que nos da la definición 4.1 la más operativa de cara a posteriores desarrollos.

En definitiva, como apuntamos en la introducción, una variable aleatoria traslada una probabilidad de los sucesos de Ω a ciertos subconjuntos de $\mathbb{R}$, entre los que están, al menos, todos los tipos de intervalos, hecho que acabamos de comprobar, así como cada uno de los puntos de $\mathbb{R}$; en efecto, para cualquier x real,

$$[X = x] = [X \leq x] - [X < x] ,$$

de donde deducimos que

$$p[X = x] = p[X \leq x] - p[X < x] .$$

Serán también probabilizables todos los conjuntos de números reales que se puedan escribir como uniones, intersecciones o complementarios de intervalos, cuyas probabilidades se obtendrán a partir de las de los intervalos, utilizando las reglas del cálculo probabilístico que hemos estudiado en los capítulos anteriores.

4.2 σ-álgebra de Borel. Probabilidad inducida por una variable aleatoria

El último comentario plantea la necesidad de delimitar su alcance, aunque el interés sea puramente formal. En otras palabras, dado un espacio probabilístico (Ω, α, p) y una variable aleatoria X, ¿para qué subconjunto B de la recta nos podemos plantear calcular la probabilidad de que X tome valores en B? Esto es, ¿cuándo $X^{-1}(B)$ es un suceso de α?

La respuesta, que no demostraremos, es que B debe pertenecer a una σ-álgebra de subconjuntos de $\mathbb{R}$, denominada de **Borel**, que definimos a continuación.

Definición 4.2 *La σ-álgebra de Borel, β, es la σ-álgebra* más pequeña *que contiene a los intervalos de $\mathbb{R}$, entendiendo la expresión* más pequeña *en el sentido de que toda σ-álgebra a la que pertenezcan los intervalos contiene a la de Borel.*

Por tanto, la σ-álgebra de Borel contendrá, en principio, todos los tipos de intervalos de $\mathbb{R}$, es decir:

$$(a,b), \quad (a,b], \quad [a,b), \quad [a,b] \,,$$

$$(-\infty,b), \quad (a,\infty), \quad [a,\infty) \quad \text{y} \quad (-\infty,b] \,.$$

Asimismo, contendrá todas las uniones e intersecciones, a lo sumo de un número infinito numerable de intervalos, y los complementarios de éstas. Es por ello que β contiene a todos los sucesos elementales puesto que

$$\{a\} = (-\infty, a] \cap [a, \infty) \,.$$

No son éstos todos los subconjuntos de Borel, pero sí son aquellos de los cuales nos interesará hallar su probabilidad. El hecho de que los sucesos elementales pertenezcan a β implica que esta σ-álgebra contiene, al menos, todos los sucesos formados por un número finito o infinito numerable de sucesos elementales. Ahora bien, existen algunos subconjuntos de $\mathbb{R}$ que no están en β; podríamos decir que están *casi todos* los subconjuntos de $\mathbb{R}$. La σ-álgebra de Borel no es, por tanto, $\mathcal{P}(\mathbb{R})$. De todas maneras, el problema es de formalización matemática y sobrepasa los planteamientos de este libro. Lo único que debemos tener en cuenta es que nos interesa trabajar con una σ-álgebra que permita calcular las probabilidades inducidas: esa σ-álgebra es β.

La definición 4.1 y la proposición 4.2 caracterizan una variable aleatoria como una aplicación de Ω en $\mathbb{R}$ en la que la contraimagen de cualquier intervalo de $\mathbb{R}$ pertenece a la σ-álgebra α. Las consideraciones anteriores sobre la σ-álgebra β nos permiten intuir una definición equivalente de variable aleatoria a partir de los subconjuntos de Borel. Así, *una aplicación $X : \Omega \to \mathbb{R}$ que verifique que*

$$X^{-1}(B) \in \alpha \quad \text{para todo } B \in \beta \,,$$

será una variable aleatoria. La equivalencia de esta condición con las planteadas en la definición y proposición antes citadas no es de fácil demostración, por lo cual

la omitiremos. Por otra parte, advertimos al lector de la poca operatividad de la definición alternativa aquí enunciada.

Todo lo anterior permite formalizar el concepto de probabilidad inducida por una variable aleatoria, concepto éste que posibilita, ahora con rigor, trasladar la probabilidad de un subconjunto de Ω a un subconjunto de la recta real. Así, *dados un espacio probabilístico* (Ω, α, p) *y una variable aleatoria* X, *podemos definir una probabilidad sobre los subconjuntos de Borel de la recta real, que se denomina* **probabilidad inducida** *por* X, *como la aplicación*

$$p_X : \beta \to \mathbb{R} ,$$

tal que, para todo B *perteneciente a* β,

$$p_X(B) = p(X^{-1}(B)) .$$

Disponemos, entonces, de un nuevo espacio probabilístico, que será el indicado para estudiar la característica numérica dada por la variable aleatoria X. Dicho espacio probabilístico va a ser $(\mathbb{R}, \beta, p_X)$, donde $\mathbb{R}$ es la recta real, β la σ-álgebra de Borel y p_X la probabilidad inducida por X. La siguiente proposición demuestra que p_X es realmente una probabilidad sobre $(\mathbb{R}, \beta)$.

Proposición 4.3 *Sea un espacio probabilístico* (Ω, α, p) *y sea* X *una variable aleatoria definida en él. La aplicación* $p_X : \beta \to \mathbb{R}$, *tal que*

$$p_X(B) = p(X^{-1}(B)) \quad \textit{para todo} B \in \beta ,$$

es una probabilidad sobre el espacio probabilizable $(\mathbb{R}, \beta)$.

Demostración. Comprobemos que verifica los tres axiomas de KOLMOGOROV.

Como para cualquier $B \in \beta$, $p_X(B) = p(X^{-1}(B))$, donde p es una probabilidad, se cumplirá que $p_X(B) \geq 0$.

Por otra parte, $p_X(\mathbb{R}) = p(X^{-1}(\mathbb{R})) = p(\Omega) = 1$.

Finalmente, si $\{B_i\}_{i=1}^{\infty}$ es una colección de sucesos de Borel disjuntos dos a dos,

$$p_X\left(\bigcup_{i=1}^{\infty} B_i\right) = p\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = p\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right) .$$

Ahora bien, como los B_i son disjuntos dos a dos, entonces

$$X^{-1}(B_i) \cap X^{-1}(B_j) = X^{-1}(B_i \cap B_j) = X^{-1}(\emptyset) = \emptyset,$$

con lo que los $\{X^{-1}(B_i)\}_{i=1}^{\infty}$ también lo son. En consecuencia,

$$p_X\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} p(X^{-1}(B_i)) = \sum_{i=1}^{\infty} p_X(B_i) .$$

Por tanto, p_X es una probabilidad sobre $(\mathbb{R}, \beta)$. ♣

4.3 Función de distribución de una variable aleatoria. Propiedades

Dado un espacio probabilístico, (Ω, α, p), y una variable aleatoria, X, su probabilidad inducida, p_X, nos permite trasladar el estudio de las probabilidades sobre los sucesos de Ω al de las probabilidades sobre subconjuntos reales. No obstante, la probabilidad inducida, como función que aplica conjuntos en números reales, resulta difícil de manejar.

En este apartado vamos a definir y estudiar la función de distribución de una variable aleatoria, que es una función numérica (de $\mathbb{R}$ en $\mathbb{R}$) y que permite, también, calcular las probabilidades sobre X.

Definición 4.3 *Sea (Ω, α, p) un espacio probabilístico y X una variable aleatoria sobre Ω. Llamamos función de distribución de X a la función*

$$F_X : I\!R \longrightarrow I\!R \,,$$

donde

$$F_X(x) = p[X \leq x] \,,$$

para todo número real x.

Adviértase que la función de distribución se puede definir, tanto en términos de la probabilidad, p, del espacio (Ω, α), como de la probabilidad inducida, p_X, de $(\mathbb{R}, \beta)$; en efecto:

$$F_X(x) = p[X \leq x] = p[X^{-1}((-\infty, x])] = p_X((-\infty, x]) \,.$$

En este sentido, $F_X(x)$ es la probabilidad inducida por la variable X en el intervalo $(-\infty, x]$, es decir, la probabilidad inducida *hasta* x.

El subíndice X de la función de distribución puede suprimirse, si ello no induce a confusión.

Ejemplo 4.4. Se lanza una moneda perfecta dos veces. Sea X la variable aleatoria *número de caras obtenidas* (esta variable ha sido estudiada en el ejemplo 4.1). Vamos a calcular su función de distribución.

El espacio muestral es

$$\Omega = \{(c,c), (c,z), (z,c), (z,z)\},$$

donde por c entendemos el resultado *obtener cara* y por z *obtener cruz*. La variable aleatoria X es una aplicación tal que $X[(c,c)] = 2$, $X[(c,z)] = X[(z,c)] = 1$, $X[(z,z)] = 0$. Como, si Ω es finito, se considera, por defecto, que todos sus subconjuntos son sucesos, cualquier aplicación de Ω en $\mathbb{R}$ es variable aleatoria, y, por tanto, lo es X.

Calculemos $F_X(x)$ para diferentes números reales x. Por ejemplo,

$$\begin{aligned} F_X(-1) &= p[X \leq -1] = p(\emptyset) = 0 \,, \\ F_X(0.5) &= p[X \leq 0.5] = p[X = 0] = p[(z,z)] = 1/4 \,, \\ F_X(1.5) &= p[X \leq 1.5] = p[(z,z),(c,z),(z,c)] = 3/4 \quad \text{o} \\ F_X(2.5) &= p[X \leq 2.5] = p(\Omega) = 1 \,. \end{aligned}$$

Formalmente, hay que calcular la función de distribución para cada número real x, igual que hicimos para los cuatro anteriores; el lector puede comprobar que la función de distribución vale

$$F_X(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ 1/4 & \text{si} \quad 0 \leq x < 1 \\ 3/4 & \text{si} \quad 1 \leq x < 2 \\ 1 & \text{si} \quad x \geq 2 \end{cases}$$

Obsérvese que las desigualdades $\leq$ y $<$ no están puestas al azar y, por tanto, si cambiamos alguna de ellas, cambia la función. En concreto, para $x = 2$, $F_X(2)$ no vale $3/4$, sino 1.

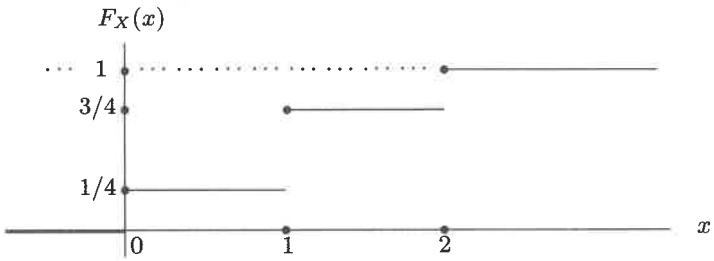


FIG. 4.3.

La figura 4.3 proporciona la representación gráfica de esta función de distribución. ◁

Las propiedades más importantes de la función de distribución se presentan en la siguiente proposición.

Proposición 4.4 *Sea X una variable aleatoria y F su función de distribución. Entonces:*

1. $0 \leq F(x) \leq 1$ *para todo número real x.*
2. *F es monótona creciente. Es decir, si x_1 y x_2 son dos números reales tales que $x_1 < x_2$, entonces $F(x_1) \leq F(x_2)$.*
3. $\lim_{x \to \infty} F(x) = 1 \quad y \quad \lim_{x \to -\infty} F(x) = 0.$
4. *Para cualquier a y b, números reales tales que $a < b$,*

$$p[a < X \leq b] = F(b) - F(a) .$$

5. *F es continua por la derecha. Esto es, para todo número real x se cumple*

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x + \varepsilon) = F(x) .$$

Demostración. El apartado *1* es obvio, ya que $F(x) = p[X \leq x]$ es una probabilidad y, por tanto, está entre 0 y 1.

Para demostrar el apartado *2*, consideremos dos números reales x_1 y x_2 tales que $x_1 < x_2$. Si $\omega \in [X \leq x_1]$, $X(\omega) \leq x_1 < x_2$. Entonces $X(\omega) < x_2$ y, consecuentemente,

$\omega \in [X \leq x_2]$. En definitiva, $[X \leq x_1] \subset [X \leq x_2]$, con lo que $F(x_1) = p[X \leq x_1] \leq p[X \leq x_2] = F(x_2)$.

Obsérvese que la función de distribución es creciente, aunque no necesariamente el crecimiento será estricto. Por ejemplo, la función de distribución construida en el ejemplo 4.4 no era estrictamente creciente; en ciertos intervalos era constante, sin crecimientos ni decrecimientos.

La demostración de las propiedades *3* y *5* exige definir límites de sucesiones de subconjuntos, lo cual rebasa el nivel de este libro. Daremos únicamente una demostración heurística de ambas. Probemos, de esta forma, la tercera propiedad[1].

Cuando x tiende hacia ∞, el suceso $[X \leq x]$ tiende al suceso $[X < \infty] = \Omega$. Así, tendremos que

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} p[X \leq x] = p(\Omega) = 1 .$$

Análogamente, cuando x tiende hacia $-\infty$, el suceso $[X \leq x]$ tiende a $[X < -\infty] = \emptyset$, y

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} p[X \leq x] = p(\emptyset) = 0 .$$

La propiedad *4* nos indica cómo se pueden calcular probabilidades de ciertos intervalos a partir de la función de distribución. Para demostrarla, obsérvese que

$$(-\infty, b] = (-\infty, a] \cup (a, b] ,$$

y que, además, $(-\infty, a] \cap (a, b] = \emptyset$. De este modo, y dado que la imagen recíproca conserva las uniones e intersecciones,

$$[X \leq b] = [X \leq a] \cup [a < X \leq b] ,$$

siendo $[X \leq a]$ y $[a < X \leq b]$ disjuntos. Por las propiedades de las probabilidades,

$$p[X \leq b] = p[X \leq a] + p[a < X \leq b]$$

y, en definitiva,

$$p[a < X \leq b] = p[X \leq b] - p[X \leq a] = F(b) - F(a) .$$

Para probar la propiedad *5* basta con comprobar que

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} (F(x + \varepsilon) - F(x)) = 0 .$$

Como hemos demostrado en la propiedad *4*,

$$F(x + \varepsilon) - F(x) = p[x < X \leq x + \varepsilon] .$$

Cuando ε tiende a cero por la derecha, el suceso $[x < X \leq x + \varepsilon]$ tiende al suceso $[x < X \leq x]$, esto es, al $\emptyset$ y, por tanto, su probabilidad tenderá a la del suceso vacío, que es cero. En definitiva,

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} (F(x + \varepsilon) - F(x)) = p(\emptyset) = 0 .$$

En las demostraciones de las propiedades *3* y *5*, se utiliza el concepto de límite de sucesiones de subconjuntos, sin haber sido definido previamente. Además, se afirma implícitamente que cuando una sucesión de sucesos tiende a otro, el límite de sus probabilidades es la probabilidad de este último, es decir, la probabilidad del límite. La

1. A menudo, con el fin de simplificar la notación, designaremos $\lim_{x \to \infty} F(x)$ y $\lim_{x \to -\infty} F(x)$ por $F(+\infty)$ y $F(-\infty)$, respectivamente.

probabilidad sería, en cierto modo, *continua*. Este resultado es cierto, pero de complicada demostración. ♣

OBSERVACIONES

1) La función de distribución de una variable aleatoria es la función que describe la acumulación de probabilidad provocada por la variable a lo largo de la recta real. En este sentido, los tramos de la misma con fuertes crecimientos indican zonas de valores de la variable que aportan gran cantidad de probabilidad, mientras que los tramos *planos* ponen en evidencia zonas en las que no hay aporte alguno. Como demostraremos en la tercera observación, los *saltos* de la función de distribución muestran aportaciones de probabilidad cuantitativamente significativas de los puntos en los que existe discontinuidad. Así, la función representada en la figura 4.4 indica aportes nulos hasta el 1, significativos en el 1 y el 2, continuos e intensos entre el 1 y el 2 y más suaves a partir del 2.

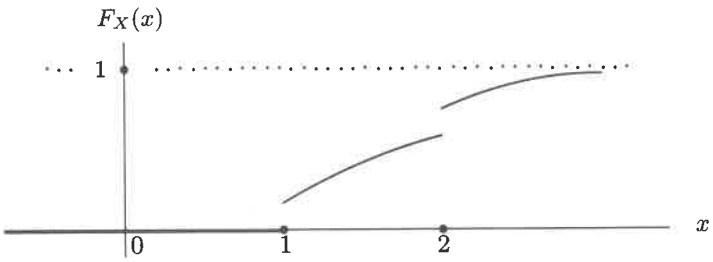


FIG. 4.4.

Como otros conceptos del Cálculo de Probabilidades, la función de distribución tiene su antecedente en la frecuencia relativa acumulada de una variable estadística, que proporcionaba el porcentaje de valores observados menores o iguales a uno dado.

2) La función de distribución no es siempre una función continua, ya que no tiene por qué ser continua por la izquierda, como veremos utilizando un razonamiento heurístico. Si ε es un número positivo, por la propiedad *4* obtenemos:

$$F(x) - F(x-\varepsilon) = p[x-\varepsilon < X \leq x] \,.$$

Pero, cuando ε tiende a cero (con $\varepsilon > 0$), el suceso $[x-\varepsilon < X \leq x]$ tiende al suceso $[X = x]$. Entonces, su probabilidad tenderá a la de dicho suceso:

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} (F(x) - F(x-\varepsilon)) = \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} p[x-\varepsilon < X \leq x] = p[X = x] \,.$$

En definitiva[2],

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x - \varepsilon) = F(x) - p[X = x] .$$

Como puede observarse, la función de distribución no tiene que ser continua por la izquierda (hecho que se había puesto ya en evidencia en el ejemplo 4.4). Si en un punto, x_0, lo fuera, F sería continua en ese punto y $p[X = x_0]$ sería igual a cero. Por tanto, las discontinuidades de F nos indican los puntos en los que la probabilidad es distinta de cero, y el *salto* de F en uno de esos puntos es, precisamente, la probabilidad del mismo.

3) La propiedad *4* nos permite hallar, a partir de la función de distribución, probabilidades de intervalos de la forma $(a, b]$, esto es, abiertos por la izquierda y cerrados por la derecha. Pero si tenemos en cuenta, como acabamos de comentar, que para cualquier número real x,

$$p[X = x] = F(x) - F(x-),$$

podemos obtener, mediante la función de distribución, la probabilidad de un intervalo cualquiera. Por ejemplo, la del intervalo cerrado $[a, b]$ sería:

$$p[a \leq X \leq b] = p[X = a] + p[a < X \leq b] .$$

Como $p[X = a] = F(a) - F(a-)$ y $p[a < X \leq b] = F(b) - F(a)$, entonces

$$p[a \leq X \leq b] = F(b) - F(a-) .$$

Análogamente, puede comprobar el lector que

$$p[a < X < b] = F(b-) - F(a)$$

y

$$p[a \leq X < b] = F(b-) - F(a-) .$$

Con todo, como veremos más adelante, no es usual que trabajemos con la función de distribución de una variable aleatoria, por lo que, salvo en ciertas situaciones, estos resultados tienen un interés más teórico que práctico.

Cuando se estudia una variable aleatoria, la aplicación de Ω en $\mathbb{R}$ no suele ser conocida (muchas veces ni se plantea cuál es Ω). Sus probabilidades se estudian, no a partir de la probabilidad, p, dada sobre los sucesos de Ω, sino asignando directamente una función de distribución, F, a la variable aleatoria X. De hecho, se propone una función (en general, de distribución, aunque más adelante se verá que puede ser de otro tipo para variables especiales) que describa cómo se *asienta* la probabilidad sobre un conjunto *discreto* o *continuo* de valores. En este sentido,

2. A partir de ahora, y para simplificar la notación, designaremos por $F(x+)$ y $F(x-)$ a

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x + \varepsilon) \quad \text{y} \quad \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x - \varepsilon) ,$$

respectivamente.

será más habitual hablar de **distribución univariante de probabilidad** que de variable aleatoria unidimensional.

El siguiente resultado describe el *aspecto* que han de tener las funciones de distribución, es decir, qué condiciones debe cumplir una función para que sea función de distribución. Como es obvio, el que una función verifique estas condiciones no va a asegurar que explique satisfactoriamente el comportamiento aleatorio de una variable concreta.

Proposición 4.5 *Sea G una función de $\mathbb{R}$ en $\mathbb{R}$. Entonces, G es la función de distribución de una variable aleatoria si y sólo si:*

1. *G es monótona creciente.*
2. *$\lim_{x\to\infty} G(x) = 1 \quad y \quad \lim_{x\to-\infty} G(x) = 0.$*
3. *G es continua por la derecha.*

Demostración. Como sabemos que toda función de distribución cumple estas propiedades, sólo tenemos que probar que la condición es suficiente, es decir, que dada una función con dichas propiedades, existe una variable aleatoria de la cual es función de distribución.

Sea una función $G : \mathbb{R} \to \mathbb{R}$ que cumple las tres condiciones del enunciado, y sea un espacio probabilístico (Ω, α, p), donde $\Omega = \mathbb{R}$, $\alpha = \beta$ y la aplicación p se define a partir de la familia de intervalos $\{(-\infty, x]\}_{x\in\mathbb{R}}$ como $p[(-\infty, x]] = G(x)$. El lector puede comprobar que dicha aplicación es, realmente, una probabilidad.

Consideremos la aplicación identidad,

$$X : \mathbb{R} \longrightarrow \mathbb{R}, \text{ tal que } X(a) = a \quad \text{para todo} \quad a \in \mathbb{R}\,.$$

Trivialmente, X es una variable aleatoria, cuya función de distribución será:

$$F_X(x) = p[X \leq x] = p[X^{-1}((-\infty, x])] = p[(-\infty, x]] = G(x)\,.$$

Luego G es una función de distribución porque coincide con F_X que lo es. ♣

Ejemplo 4.5. Sea X la variable aleatoria que mide *el tiempo, en minutos, que tarda en ser atendido un cliente* en la caja de un hipermercado. Podemos asignar a X la siguiente función de distribución:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

La gráfica de dicha función es la mostrada en la figura 4.5.

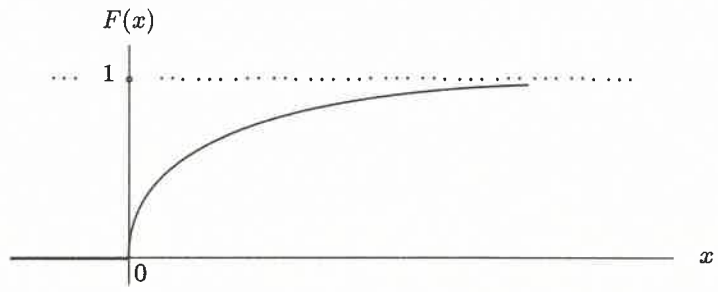


FIG. 4.5.

Se demuestra, sin dificultad, que dicha función es una función de distribución; en efecto:

1. F es una función monótona creciente. Evidentemente, para los números negativos es constante (vale cero).
   Si $x_1 < 0 < x_2$,
   $$F(x_1) = 0 < 1 - e^{-x_2} = F(x_2) .$$
   Si $0 < x_1 < x_2$, $e^{-x_1} > e^{-x_2}$ y, por tanto,
   $$F(x_1) = 1 - e^{-x_1} < 1 - e^{-x_2} = F(x_2) .$$
2. Obviamente, $F(-\infty) = 0$. Por otra parte,
   $$\lim_{x\to\infty} F(x) = \lim_{x\to\infty} (1 - e^{-x}) = 1 - \lim_{x\to\infty} e^{-x} = 1 .$$
3. F es continua por la derecha. Más aún, es continua en toda la recta. El lector puede comprobar que en el único punto conflictivo, $x = 0$, coinciden los límites por la derecha y por la izquierda, siendo su valor igual a $F(0)$.

En definitiva, F es una función de distribución en virtud de la proposición 4.5. Advertirá el lector que la simple observación de la gráfica pone en evidencia las tres propiedades que caracterizan a una función de distribución.

Con ella pueden calcularse probabilidades para la variable X. Por ejemplo, la probabilidad de que una persona se encuentre en la cola más de un minuto, pero no más de dos, vale

$$p[1 < X \leq 2] = F(2) - F(1) = (1 - e^{-2}) - (1 - e^{-1}) = e^{-1} - e^{-2} = 0.23 ,$$

o bien, la probabilidad de que tarde más de tres minutos en ser atendida valdrá

$$p[X > 3] = 1 - p[X \leq 3] = 1 - F(3) = 1 - (1 - e^{-3}) = e^{-3} = 0.05 .$$

Esta función de distribución se asocia, con frecuencia, a variables aleatorias que miden lo que se denomina *tiempos de espera*, como es la anterior, porque responde bien al comportamiento probabilístico de este tipo de fenómenos. Como puede observarse, dado que F crece rápidamente al principio, la probabilidad de que la espera sea pequeña es alta. En cambio, como el crecimiento de F se suaviza para los valores altos de la variable, los tiempos de espera grandes son muy poco probables.

Puede buscarse una función de distribución parecida a ésta si, por ejemplo, la cajera del hipermecado es muy lenta en el servicio. Puede comprobar el lector que la función

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - ae^{-ax} & x \geq 0 \end{cases}$$

donde a es un número real positivo fijo, es una función de distribución. Cuanto más grande sea el valor de a, mayor es el crecimiento inicial de F (la cajera será más rápida) y, recíprocamente, valores pequeños de a se utilizarían para cajeras muy lentas (F crece más despacio). Puede realizar el lector diversos cálculos de probabilidades para distintos valores de a. Por ejemplo, la probabilidad de que el tiempo de espera no rebase los dos minutos, $p[X \leq 2]$, vale 0.82 para $a = 0.5$, vale 0.86 para $a = 1$ y 0.96 para $a = 2$. ◁

Adviértase que dos o más variables pueden tener la misma distribución. Por ejemplo, X, *número de caras obtenidas en el lanzamiento de una moneda*, e Y, *número de resultados pares obtenidos al tirar un dado*, son variables distintas, pero, si se supone que tanto la moneda como el dado están sin cargar, tienen la misma distribución. Así,

$$X : \left| \begin{array}{ll} 1 & \text{si se obtiene cara} \\ 0 & \text{si se obtiene cruz} \end{array} \right.$$

y

$$Y : \left| \begin{array}{ll} 1 & \text{si sale 2, 4 o 6} \\ 0 & \text{si sale 1, 3 o 5} \end{array} \right.$$

pero

$$p[X = 1] = p[Y = 1] = \frac{1}{2}$$

y

$$p[X = 0] = p[Y = 0] = \frac{1}{2}.$$

Por tanto,

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

es la función de distribución que describe la acumulación de probabilidad, tanto de X, como de Y.

Diremos que *dos o más variables aletorias están* **igual o idénticamente distribuidas** *si tienen la misma función de distribución.*

La utilización de las funciones de distribución para calcular probabilidades es, como hemos visto, muy engorrosa. Sin embargo, en la mayor parte de las situaciones prácticas, las variables aleatorias que se nos presentan son de tal forma que disponemos de otros recursos más sencillos para calcular probabilidades. Ello ocurre, en particular, cuando las variables aleatorias son, o bien de las denominadas discretas, o bien de las denominadas continuas.

4.4 Variables aleatorias discretas

Estudiaremos, en este apartado, un primer tipo de variables, conceptualmente sencillas, cuyos desarrollos no requieren conocimientos matemáticos muy avanzados.

Definición 4.4 *Sea (Ω, α, p) un espacio probabilístico y X una variable aleatoria definida en él. Decimos que X es una variable aleatoria discreta si el conjunto*

$$X(\Omega) = \{X(\omega),\ \textit{para todo}\ \omega \in \Omega\}$$

es finito o infinito numerable[3].

Desarrollaremos toda esta sección suponiendo que $X(\Omega)$ es infinito numerable, esto es,

$$X(\Omega) = \{x_1, x_2, \ldots, x_n, \ldots\}\,,$$

donde $x_1 < x_2 < \cdots < x_n < \cdots$. El análisis es similar si $X(\Omega)$ es finito.

Las características de este tipo de variables nos van a permitir definir una nueva función más operativa que la función de distribución.

Definición 4.5 *Sea X una variable aleatoria discreta. Su función de probabilidad es una aplicación,*

$$f_X : \mathbb{R} \to \mathbb{R}\,,$$

definida como

$$f_X(x) = p[X = x] \qquad \textit{para todo}\ x \in \mathbb{R}\,.$$

Advierta el lector que, si $X(\Omega)$ es el conjunto $\{x_1, x_2, \ldots, x_n, \ldots\}$, la función de probabilidad toma los valores

$$f_X(x_i) = p[X = x_i] \qquad \text{para } i = 1, 2, \ldots, n, \ldots$$

y es igual a cero para el resto de los números reales.

Buscando una mejor visualización de la distribución, la representación de la función de probabilidad se realiza sólo para los puntos en los que toma valores no nulos.

3. Recordemos que un conjunto es numerable si existe una aplicación biyectiva entre él y los naturales, es decir, si es del mismo *tamaño* que el conjunto de los números naturales. Por ejemplo, el conjunto de números racionales, $\mathbb{Q}$, es numerable, mientras que el conjunto $\mathbb{R}$ de los números reales no lo es. No obstante, existen conjuntos numerables de aspecto muy diferente. A nosotros sólo nos interesa estudiar variables discretas tales que $X(\Omega)$, si es numerable, tenga el aspecto de los números naturales. En concreto, nos interesa que, dado un punto $x_i \in X(\Omega)$, exista su inmediato anterior, x_{i-1} (salvo si es el primero, x_1), y su inmediato posterior, x_{i+1}, con el orden usual entre números reales. Los racionales, entre otros, no cumplen esta condición. Por ejemplo, no existe un número racional que sea el inmediato anterior o el inmediato posterior al racional 2.5. A partir de ahora, si una variable discreta toma una infinidad numerable de valores, supondremos que $X(\Omega)$ cumple la condición anterior, es decir, tiene el aspecto de los naturales.

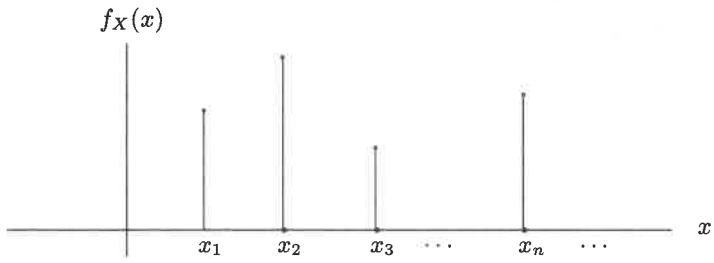


FIG. 4.6.

La función de probabilidad, en cuya notación omitiremos el subíndice X siempre que no se preste a confusión, permite calcular las probabilidades para una variable aleatoria discreta.

Proposición 4.6 *Sea X una variable aleatoria y f su función de probabilidad. Si B es un subconjunto de $\mathbb{R}$,*

$$p[X \in B] = \sum_{x_i \in B} f(x_i) .$$

Demostración. En efecto, como X toma como valores sólo los x_i, para cualquier $B \subset \mathbb{R}$,

$$[X \in B] = \bigcup_{x_i \in B} [X = x_i] ,$$

unión que será de una colección finita o infinita numerable. Además, si $i \neq j$

$$[X = x_i] \cap [X = x_j] = \emptyset ,$$

ya que, al ser X una aplicación, no puede haber ningún $\omega \in \Omega$ tal que $X(\omega) = x_i$ y $X(\omega) = x_j$.

Por tanto, hemos escrito $[X \in B]$ como unión de una colección finita o infinita numerable de sucesos disjuntos dos a dos, con lo que

$$p[X \in B] = \sum_{x_i \in B} p[X = x_i] = \sum_{x_i \in B} f(x_i) ,$$

siendo f la función de probabilidad de X. ♣

La función de distribución se calcula a partir de la función de probabilidad, sin más que aplicar la proposición anterior. Así,

$$F(x) = p[X \leq x] = \sum_{x_i \leq x} p[X = x_i] = \sum_{x_i \leq x} f(x_i) .$$

Obsérvese que la función de distribución de una variable aleatoria discreta tiene un aspecto característico. En efecto, si $x_k \leq x < y < x_{k+1}$,

$$F(x) = p[X \leq x] = \sum_{x_i \leq x} p[X = x_i] = \sum_{i=1}^{k} p[X = x_i]$$

y

$$F(y) = p[X \leq y] = \sum_{x_i \leq y} p[X = x_i] = \sum_{i=1}^{k} p[X = x_i] \ ,$$

y, por tanto, $F(x) = F(y)$. En definitiva, F es constante en los intervalos $[x_k, x_{k+1})$.

Además, F tendrá discontinuidades en los puntos x_k $(k = 1, 2, \ldots)$, como vimos en la observación **2** de la sección 4.3, o como se puede observar directamente. La discontinuidad, es decir, *el salto*, en un x_k, es por la izquierda, y vale

$$F(x_k) - F(x_k-) = p[X = x_k] = f(x_k) \ .$$

Resumiendo, la función de distribución de una variable aleatoria discreta es una función (finita o numerablemente) escalonada. Este resultado, junto con el hecho de que las discontinuidades son por la izquierda, caracteriza la función de distribución de una variable aleatoria discreta cualquiera. La figura 4.7 ilustra su aspecto.

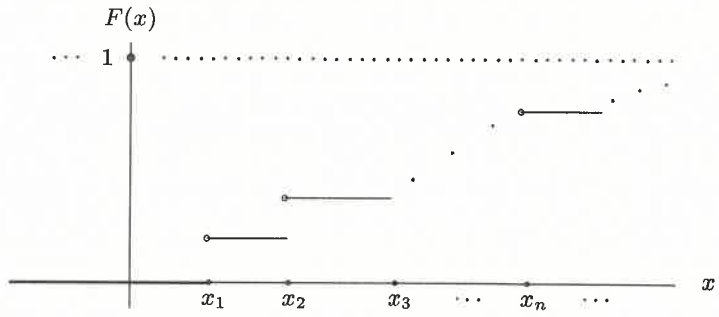


FIG. 4.7.

Aunque no sea un resultado de utilidad práctica, se puede obtener a partir de la función de distribución la función de probabilidad. Como la función de distribución es constante entre dos valores consecutivos de la variable,

$$F(x_k-) = F(x_{k-1}) \ ,$$

y el resultado anterior se convierte en

$$f(x_k) = p[X = x_k] = F(x_k) - F(x_{k-1}) \ , \quad \text{para} \quad k = 2, 3, \ldots$$

Obviamente, $f(x_1) = F(x_1)$. Por tanto, la función de probabilidad en un punto es el *salto* que da la función de distribución en dicho punto.

La siguiente proposición resume la relación, ya demostrada, entre ambas funciones.

Proposición 4.7 *Sea X una variable aleatoria discreta y sean F y f su función de distribución y de probabilidad, respectivamente. Entonces,*

1. $F(x) = \sum_{x_i \leq x} f(x_i)$ *para todo* $x \in \mathbb{R}$.
2. $f(x_k) = F(x_k) - F(x_{k-1})$ *para* $k = 2, 3, \ldots$ *y* $f(x_1) = F(x_1)$.

Ejemplo 4.6. Vamos a hallar las funciones de distribución y de probabilidad de la variable X, *número total de bolas blancas obtenidas* al hacer dos extracciones con reemplazamiento en una urna compuesta por tres bolas, una blanca, una roja y una negra (dicha variable se ha estudiado en el ejemplo 4.2).

La función de distribución de una variable aleatoria está definida para todo número real, por tanto, deberíamos calcular $p[X \leq x]$ para todo $x \in \mathbb{R}$. Así, por ejemplo, si $0 \leq x < 1$,

$$F_X(x) = p[X \leq x] = p[X \leq 0] = p[(r,r),(r,n),(n,r),(n,n)] = \frac{4}{9}.$$

Haciendo los cálculos para el resto de valores de x, obtenemos la siguiente función de distribución:

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 4/9 & \text{si } 0 \leq x < 1 \\ 8/9 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

La gráfica de esta función de distribución es la mostrada en la figura 4.8.

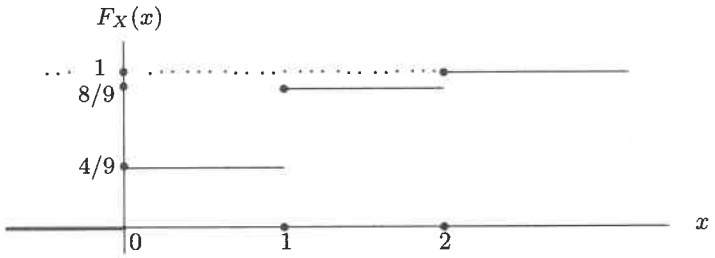


FIG. 4.8.

Es evidente que esta variable es discreta pues, como ya hemos visto, $X(\Omega) = \{0, 1, 2\}$. Calculemos su función de probabilidad para los valores de la variable:

$$\begin{aligned} p[X = 0] &= p[(r,r),(r,n),(n,r),(n,n)] = \frac{4}{9}, \\ p[X = 1] &= p[(b,r),(b,n),(r,b),(n,b)] = \frac{4}{9} \quad \text{y} \\ p[X = 2] &= p[(b,b)] = \frac{1}{9}. \end{aligned}$$

Resumiendo, la función de probabilidad es

$$f_X(x) = \begin{cases} 4/9 & x = 0 \\ 4/9 & x = 1 \\ 1/9 & x = 2 \\ 0 & \text{en el resto,} \end{cases}$$

y su representación gráfica es la mostrada en la figura 4.9.

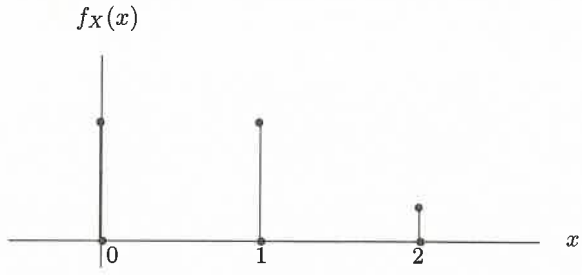


FIG. 4.9.

Nótese que podríamos haber calculado la función de probabilidad a partir de la de distribución. Así,

$$\begin{aligned} p[X=0] &= F_X(0) - F_X(0-) = \frac{4}{9} - 0 = \frac{4}{9}\,, \\ p[X=1] &= F_X(1) - F_X(0) = \frac{8}{9} - \frac{4}{9} = \frac{4}{9} \quad \text{y} \\ p[X=2] &= F_X(2) - F_X(1) = 1 - \frac{8}{9} = \frac{1}{9}\,. \end{aligned}$$

De este modo, hemos corroborado que el *salto* de la función de distribución en cada uno de los valores de la variable es, precisamente, el valor de la función de probabilidad en dicho punto. ◁

Como puede observarse, la función de probabilidad permite realizar cualquier cálculo probabilístico para una variable aleatoria discreta. En muchos casos, no se construye a partir del espacio probabilístico de partida, sino que se asigna directamente a una variable aleatoria. Resulta, por ello, especialmente importante estudiar el aspecto de las funciones de probabilidad.

Proposición 4.8

1. *Sea f la función de probabilidad de una variable aleatoria discreta, X, que toma los valores $x_1, \ldots, x_n, \ldots$ Entonces,*

   (a) $0 \leq f(x) \leq 1$ *para todo* $x \in \mathbb{R}$.

   (b) $\sum_{i=1}^{\infty} f(x_i) = 1$.

2. *Además, si $g : \mathbb{R} \to \mathbb{R}$ es una función que se anula salvo para un conjunto finito o infinito numerable, $\{x_1, \ldots, x_n, \ldots\}$, y cumple*

   (a) $0 \leq g(x_i) \leq 1$ *para todo* x_i,

   (b) $\sum_{i=1}^{\infty} g(x_i) = 1$,

   entonces, g es la función de probabilidad de una variable aleatoria discreta.

Demostración. Probemos el primer apartado. Evidentemente, $0 \leq f(x) \leq 1$ puesto que, por definición, $f(x)$ es una probabilidad. Además, como $\Omega = \cup_{i=1}^{\infty}[X = x_i]$ y dichos sucesos son disjuntos dos a dos,

$$\sum_{i=1}^{\infty} f(x_i) = \sum_{i=1}^{\infty} p[X = x_i] = p\left[\bigcup_{i=1}^{\infty}[X = x_i]\right] = p(\Omega) = 1 .$$

Probemos, ahora, el apartado *2*. Sea $G(x) = \sum_{x_i \leq x} g(x_i)$. Puede comprobarse que G es monótona creciente, continua por la derecha y que $G(\infty) = 1$ y $G(-\infty) = 0$. G es, por tanto, la función de distribución de una variable aleatoria discreta que toma los valores $\{x_1, \ldots, x_n, \ldots\}$, siendo g su correspondiente función de probabilidad. ♣

Ejemplo 4.7. Un fabricante de tornillos afirma que el 5 % son defectuosos. Disponemos de una caja de 100 unidades. Estudiemos la variable aleatoria *número de extracciones hasta encontrar el primer tornillo defectuoso*. Parece evidente que los resultados dependerán de que devolvamos, o no, a la caja el tornillo extraído.

CASO 1. Las extracciones se realizan con reemplazamiento.

La variable toma como valores posibles todos los números naturales, con la siguiente función de probabilidad:

$$f(1) = p[X = 1] = p(D_1) = 0.05 ,$$
$$f(2) = p[X = 2] = p(\overline{D}_1 D_2) = p(\overline{D}_1)p(D_2) = 0.95 \cdot 0.05 ,$$

donde D_i indica el suceso *obtener un tornillo defectuoso en la extracción i-ésima*, y, en general,

$$f(k) = p[X = k] = p(\overline{D}_1 \ldots \overline{D}_{k-1} D_k) = 0.95^{k-1} \cdot 0.05 , \quad k = 1, 2, \ldots$$

Comprobemos que f es una función de probabilidad. Obviamente, $0 \leq f(k) \leq 1$ y, además,

$$\sum_{k=1}^{\infty} f(k) = \sum_{k=1}^{\infty} 0.05 \cdot 0.95^{k-1} = 0.05 \cdot \sum_{k=1}^{\infty} 0.95^{k-1} = 0.05 \cdot \frac{1}{1 - 0.95} = 1 .$$

La representación gráfica de $f(x)$ es la mostrada en la figura 4.10.

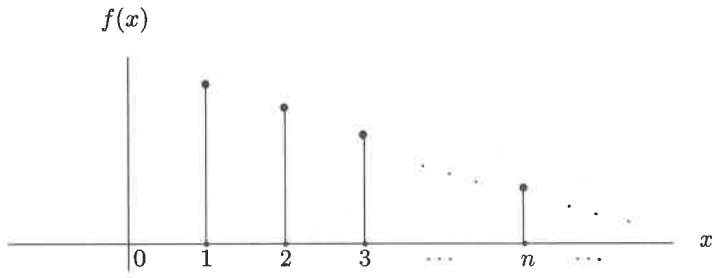


FIG. 4.10.

El lector puede comprobar que su función de distribución es:

$$F(x) = p[X \leq x] = \begin{cases} 0 & x < 1 \\ 0.05 & 1 \leq x < 2 \\ 0.05 + 0.05 \cdot 0.95 & 2 \leq x < 3 \\ \vdots & \vdots \\ 0.05 + 0.05 \cdot 0.95 + \cdots + 0.05 \cdot 0.95^{k-1} & k \leq x < k+1 \\ \vdots & \vdots \end{cases}$$

Teniendo en cuenta que cada valor de la función de distribución es la suma de una progresión geométrica limitada, podemos dar la siguiente expresión general:

$$F(x) = \begin{cases} 1 - 0.95^{k-1} & k \leq x < k+1, \quad k = 1, 2, \ldots \\ 0 & x < 1 \end{cases}$$

cuya representación gráfica es la mostrada en la figura 4.11.

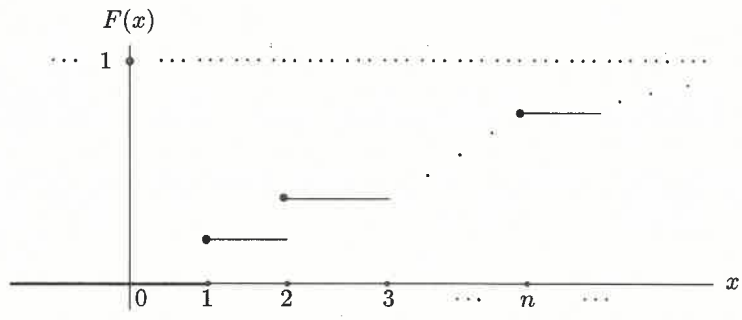


FIG. 4.11.

A partir de la función de probabilidad, podemos calcular la probabilidad de cualquier suceso. Así, la probabilidad de *realizar al menos* 10 *extracciones hasta encontrar un tornillo defectuoso* será:

$$p[X \geq 10] = 1 - p[X < 10] = 1 - \sum_{k=1}^{9} f(k) = 0.95^8 = 0.66 .$$

Asimismo, podemos calcular la probabilidad de un suceso condicionada a la ocurrencia de otro. Por ejemplo, si en las 9 primeras extracciones no ha salido un tornillo defectuoso, podemos calcular la probabilidad de que *no haya que realizar mas de* 15. Es decir

$$p[X \leq 15 \mid X \geq 10] = \frac{p[X \leq 15\,,\ X \geq 10]}{p[X \geq 10]} =$$

$$= \frac{p[10 \leq X \leq 15]}{1 - p[X \leq 9]} = \frac{\sum_{10}^{15} f(k)}{1 - \sum_{1}^{9} f(k)} = 0.26 .$$

CASO 2. Las extracciones se realizan sin reemplazamiento.

Ahora, X es una variable discreta que toma únicamente los valores $1, 2, 3, \ldots, 96$. Su función de probabilidad será:

$$f(1) = p[X = 1] = p(D_1) = \frac{5}{100}\,,$$

$$f(2) = p[X = 2] = p(\overline{D}_1 D_2) = p(\overline{D}_1)p(D_2 \mid \overline{D}_1) = \frac{95}{100} \cdot \frac{5}{99}$$

y, en general,

$$f(k) = p[X = k] = \frac{95}{100} \cdot \frac{94}{99} \cdots \frac{95 - (k-2)}{100 - (k-2)} \cdot \frac{5}{100 - (k-1)}\,.$$

Puede comprobar el lector que se trata de una función de probabilidad y obtener la función de distribución. Lógicamente, y al igual que en el caso anterior, la función de distribución es escalonada, con *saltos* cada vez menores; sin embargo, alcanza el valor uno, en concreto, a partir de $x = 96$. ◁

Tanto en los desarrollos, como en los ejemplos, hemos considerado variables aleatorias discretas que tomaban un número finito, o infinito numerable, de valores; en este último caso, suponíamos que $X(\Omega)$ presentaba un *aspecto* como el de los números naturales. El siguiente ejemplo nos plantea una situación más general, si bien de menor relevancia práctica.

Ejemplo 4.8. Sea X una variable aleatoria tal que $X(\Omega)$ es el conjunto $\mathbb{Z}$ de los números enteros. Definamos la siguiente función:

$$f(x) = \begin{cases} \frac{1}{3} \cdot 2^{-|x|} & x = \ldots, -2, -1, 0, 1, 2, \ldots \\ 0 & \text{en el resto} \end{cases}$$

Comprobemos, en primer lugar, que es realmente una función de probabilidad. Evidentemente, $f(x)$ toma valores mayores o iguales que cero. Además, teniendo en cuenta que f es simétrica respecto del cero,

$$\sum_{x \in \mathbb{Z}} f(x) = f(0) + 2 \cdot \sum_{x=1}^{\infty} f(x) = \frac{1}{3} + \frac{2}{3} = 1\,,$$

pues $\sum_{x=1}^{\infty} 2^{-x}$ es la suma de una progresión geométrica ilimitada cuyo primer término y razón valen $1/2$. La gráfica de esta función de probabilidad puede verse en la figura 4.12.

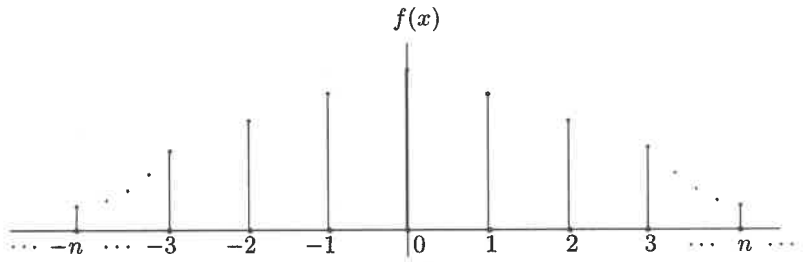


FIG. 4.12.

Con ella podemos calcular la probabilidad de cualquier suceso dado en términos de la variable aleatoria. Así, la probabilidad de que X tome valores pares y positivos será:

$$\begin{aligned} p[X \text{ sea par y positivo}] &= p[X=2]+p[X=4]+\cdots = \\ &= \frac{1}{3}\cdot 2^{-2}+\frac{1}{3}\cdot 2^{-4}+\cdots = \frac{1}{9}\,. \end{aligned}$$

O la probabilidad de que X tome valores negativos,

$$p[X<0]=\sum_{x=-\infty}^{-1}\frac{1}{3}\cdot 2^{-|x|}=\sum_{x=1}^{\infty}\frac{1}{3}\cdot 2^{-x}=\frac{1}{3}\,.$$

Por otro lado, sabiendo que X toma valores positivos, la probabilidad de que *sean pares* será:

$$\begin{aligned} p[X \text{ sea par} \mid X \text{ sea positivo}] &= \frac{p[X \text{ sea par y positivo}]}{p[X \text{ sea positivo}]} = \\ &= \frac{\sum_{k=1}^{\infty} p[X=2k]}{\sum_{k=1}^{\infty} p[X=k]} = \frac{1}{3}\,. \end{aligned}$$

El lector puede comprobar que la función de distribución es:

$$F(x)=\begin{cases} \frac{1}{3}\cdot\frac{1}{2^{k-1}} & -k\le x<-k+1 \\ \frac{2}{3} & 0\le x<1 \\ 1-\frac{1}{3}\cdot\frac{1}{2^{k}} & k\le x<k+1 \end{cases} \qquad k=1,2,\ldots.$$

Dicha función es escalonada, pero no llega nunca a alcanzar ni el cero (no hay un primer valor de la variable), ni el uno (no hay, tampoco, un último valor). ◁

4.5 Variables aleatorias continuas

El segundo tipo de variables aleatorias de tratamiento aceptablemente sencillo es el que se presenta con el nombre de variables absolutamente continuas, y que nosotros denominaremos, simplemente, continuas.

Definición 4.6 *Sea X una variable aleatoria y F_X su función de distribución. Decimos que X es una variable aleatoria absolutamente continua si*

1. *F_X es una función continua.*
2. *Su derivada, F'_X, existe y es continua salvo, a lo sumo, en un número finito de puntos.*

Las variables aleatorias continuas aparecen cuando se estudian características numéricas que pueden tomar una infinidad no numerable de valores (todos los de la recta real o los de un intervalo de la misma). En ellas, pequeñas variaciones de x producen pequeñas variaciones de $F_X(x)$, de forma que la probabilidad se

acumula suavemente a lo largo de todo el campo de variación y no sólo en unos cuantos puntos. De hecho, el nombre con que se designa a estas variables hace referencia clara al crecimiento suave, sin *saltos*, de F_X, exigencia que se satisface con la primera condición de la definición; la segunda, es una condición requerida para hacer operativa la variable.

Al igual que para las variables discretas, existe una nueva función que va a facilitar el cálculo de las probabilidades en X.

Definición 4.7 *Sea X una variable aleatoria continua. Su función de densidad es una aplicación*

$$f_X : \mathbb{R} \to \mathbb{R} ,$$

definida como

$$f_X(x) = F'_X(x) = \frac{d}{dx} F_X(x) ,$$

donde F_X es la función de distribución de X.

La función de densidad es una aplicación de $\mathbb{R}$ en $\mathbb{R}$, por ser la derivada de una función real de variable real; $f_X(x)$ está definida para todos los números reales x, salvo, a lo sumo, para un número finito de ellos. Más adelante analizaremos que este hecho es irrelevante para su uso. Como ya comentamos en circunstancias análogas, normalmente omitiremos los subíndices de F_X y f_X.

La siguiente proposición relaciona las funciones de distribución y de densidad de una variable aleatoria continua.

Proposición 4.9 *Sea X una variable aleatoria continua y sean F y f sus funciones de distribución y de densidad, respectivamente. Entonces,*

1. $f(x) = \frac{d}{dx} F(x)$ *para todo* $x \in \mathbb{R}$.
2. $F(x) = \int_{-\infty}^{x} f(t)dt$ *para todo* $x \in \mathbb{R}$.

Demostración. Por definición, la función de densidad se obtiene derivando la función de distribución. Asimismo, es posible obtener la función de distribución a partir de la de densidad. En efecto, F es una primitiva de f y, por tanto, si a y b son números reales cualesquiera, aplicando la fórmula de BARROW,

$$F(b) - F(a) = \int_a^b f(x)dx .$$

Tomando límites cuando a tiende a $-\infty$, como $F(-\infty) = 0$,

$$F(b) = \int_{-\infty}^{b} f(x)dx ,$$

resultado que permite obtener la función de distribución a partir de la de densidad. ♣

Nótese que ya probamos un resultado análogo al anterior para las variables aleatorias discretas (proposición 4.7). Por un lado, dicho resultado nos proporcionaba

la función de probabilidad tomando diferencias en las discontinuidades de la función de distribución; ahora, obtenemos la función de densidad tomando derivadas en la función de distribución. Por otro lado, obtuvimos la función de distribución sumando los valores de la función de probabilidad; ahora, hallamos la función de distribución integrando la función de densidad[4].

Estudiemos las propiedades de la función de densidad.

Proposición 4.10 *Sea X una variable aleatoria continua y sea f su función de densidad. Entonces:*

1. $f(x) \geq 0$ *para todo* $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. *Para cualquier a y b, números reales tales que $a < b$,*

$$p[a < X \leq b] = \int_a^b f(x)dx \, .$$

4. *Para cualquier subconjunto B de la σ-álgebra de Borel,*

$$p[X \in B] = \int_B f(x)dx \, .$$

Demostración. La primera es inmediata, por ser $f(x)$ la derivada de $F(x)$, función creciente por definición.

Para la segunda, como

$$F(b) = \int_{-\infty}^b f(x)dx \, ,$$

si tomamos límites cuando b tiende a ∞, obtendremos:

$$\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1 \, .$$

Para demostrar la tercera propiedad, tengamos en cuenta que para cualquier variable aleatoria se verifica:

$$p[a < X \leq b] = F(b) - F(a) \, .$$

Si, además, X es continua,

$$F(b) - F(a) = \int_a^b f(x)dx \, ,$$

según hemos visto anteriormente. Por tanto,

$$p[a < X \leq b] = \int_a^b f(x)dx \, .$$

El cuarto resultado es una extensión, que no demostraremos, de la propiedad anterior para cualquier subconjunto de Borel. ♣

4. El lector familiarizado con el Cálculo Diferencial e Integral sabe que la derivada y la integral pueden entenderse, en cierto sentido, como una extensión de la diferencia y la suma en el *continuo*.

OBSERVACIONES

1) Aunque se utilice la misma notación para designar ambas, existe una diferencia esencial entre la función de probabilidad de una variable discreta y la función de densidad de una variable continua. Mientras que aquélla era una probabilidad, la función de densidad en un punto no lo es (incluso, puede tomar valores mayores que uno).

De hecho, las variables continuas verifican una propiedad que aparentemente es sorprendente: si x es un número real cualquiera, $p[X = x] = 0$. En efecto, recordando las propiedades de la función de distribución, los puntos para los que la probabilidad era distinta de cero eran aquellos en los cuales la función de distribución era discontinua, en concreto por la izquierda (observación 2, sección 4.3). Como la función de distribución de una variable aleatoria continua es continua en toda la recta, cada número real tendrá probabilidad cero,

$$p[X = x] = F(x) - F(x-) = 0 \, .$$

De lo anterior no podemos deducir la imposibilidad de ocurrencia de los valores de la variable; incluso, podremos afirmar que unos (expresándolo con rigor, que unas zonas de valores) tienen más probabilidad de ocurrir que otros (que otras zonas). ¿Cómo se reconoce entonces qué valores toma una variable aleatoria continua? Si una variable es discreta, los valores que toma son aquellos en los que la función de probabilidad es distinta de cero. Sin embargo, para una variable aleatoria continua, la probabilidad de un punto es cero, tanto si es un valor que toma la variable, como si no lo es. Para una variable continua, los valores que toma la variable son aquellos en los que la función de densidad es distinta de cero, siendo *más probables* aquellos para los que la función de densidad toma valores más grandes (es decir, la función de distribución tiene pendientes más pronunciadas, acumulaciones de probabilidad mayores).

2) Puesto que los puntos tienen probabilidad cero, la probabilidad de un intervalo es la misma, tanto si contiene a uno o a los dos extremos, como si no contiene a ninguno de ellos. Por ejemplo,

$$p[a \leq X \leq b] = p[X = a] + p[a < X \leq b] = p[a < X \leq b] = \int_a^b f(x)dx \, ,$$

ya que $p[X = a] = 0$.

3) Según hemos visto, la función de densidad en un punto x no es la probabilidad de dicho punto; entonces, ¿qué expresa realmente dicha función?

Para dar respuesta a esta pregunta, tomaremos un intervalo cerrado alrededor de x de amplitud h (un entorno), es decir, $\left[x - \frac{h}{2}, x + \frac{h}{2}\right]$ y calcularemos su probabilidad. Teniendo en cuenta la observación anterior:

$$p\left[x - \frac{h}{2} \leq X \leq x + \frac{h}{2}\right] = \int_{x-\frac{h}{2}}^{x+\frac{h}{2}} f(t)dt = F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right) \, .$$

Aplicando a la función de distribución el teorema de los incrementos finitos, sabemos que existe un punto t del intervalo $(x-\frac{h}{2}, x+\frac{h}{2})$ que verifica:

$$F\left(x+\frac{h}{2}\right)-F\left(x-\frac{h}{2}\right)=h\cdot F'(t)=h\cdot f(t)\,.$$

Por tanto,

$$p\left[x-\frac{h}{2}\leq X\leq x+\frac{h}{2}\right]=h\cdot f(t)\,.$$

Dicho punto t lo podemos expresar como

$$t=\left(x-\frac{h}{2}\right)+\theta\cdot h=x+\left(\theta-\frac{1}{2}\right)\cdot h\,,$$

para un cierto θ tal que $0<\theta<1$. Entonces,

$$f\left(x+\left(\theta-\frac{1}{2}\right)\cdot h\right)=\frac{p\left[x-\frac{h}{2}\leq X\leq x+\frac{h}{2}\right]}{h}\,.$$

Teniendo en cuenta la continuidad de f (salvo, a lo sumo, en un número finito de puntos), obtendremos:

$$f(x)=\lim_{h\to 0} f\left(x+\left(\theta-\frac{1}{2}\right)\cdot h\right)=\lim_{h\to 0}\frac{p\left[x-\frac{h}{2}\leq X\leq x+\frac{h}{2}\right]}{h}\,.$$

De ello deducimos que la función de densidad en un punto x es la probabilidad por unidad de longitud en un intervalo *infinitesimal* alrededor de x. Es decir, expresa la densidad de probabilidad en torno al punto.

En la definición de variable aleatoria continua se permite que f no exista, o no sea continua, en un número finito de puntos. Veamos un ejemplo.

Ejemplo 4.9. Se considera una variable aleatoria X con función de distribución

$$F(x)=\begin{cases}0 & x<0\\ x & 0\leq x\leq 1\\ 1 & x>1\end{cases}$$

cuya representación gráfica es la de la figura 4.13.

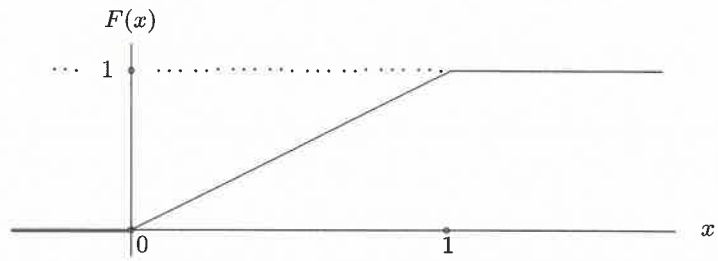


FIG. 4.13.

Obsérvese que es plana antes del cero y después del uno (no hay aportes de probabilidad) y que su crecimiento es constante en el resto (podríamos decir que todos los puntos aportan cantidades insignificativas, pero idénticas, de probabilidad).

Podemos obtener la función de densidad de X. Obviamente, F es continua y, además, salvo en $x = 0$ y $x = 1$, es derivable, y su derivada es continua. En efecto,

$$f(x) = \frac{d}{dx}F(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 < x < 1 \\ 0 & x > 1 \end{cases}$$

En 0 y 1, F no es derivable. En $x = 0$, la derivada por la izquierda vale cero, y la derivada por la derecha uno. En $x = 1$, la derivada por la izquierda de F vale uno, y por la derecha cero.

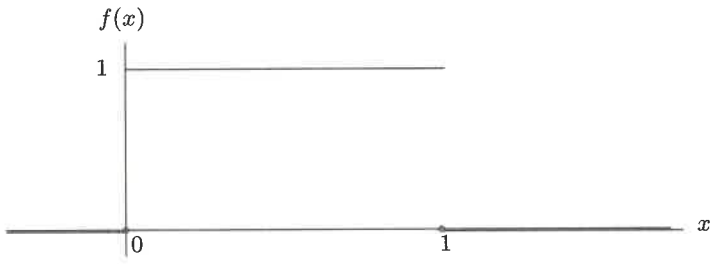


FIG. 4.14.

La función de densidad, representada en la figura 4.14, no existe para $x = 0$ y para $x = 1$. Adviértase, por otra parte, que los comentarios surgidos de la observación de la gráfica de la función de distribución se pueden trasladar a esta última.

Conocida la función de densidad, se agilizan los cálculos de las probabilidades de los sucesos que vengan expresados en términos de la variable aleatoria. Así, la probabilidad de que X *tome valores superiores a* 0.25 será:

$$p[X > 0.25] = \int_{0.25}^{1} 1 \cdot dx = 0.75 .$$

O, sabiendo que X toma valores superiores a 0.25, la probabilidad de que *sean inferiores a* 0.50 será:

$$p[X < 0.50 \mid X > 0.25] = \frac{p[0.25 < X < 0.50]}{p[X > 0.25]} = \frac{\int_{0.25}^{0.50} 1 \cdot dx}{\int_{0.25}^{1} 1 \cdot dx} = \frac{1}{3} .$$

Téngase en cuenta que todos estos cálculos también son posibles con la función de distribución. ◁

El problema planteado en el ejemplo anterior es, de hecho, irrelevante en el trabajo práctico. Recordemos que, por un lado, el cálculo de probabilidades para una variable aleatoria continua se realiza integrando la función de densidad y, por otro lado, una integral no varía si el integrando se cambia en un número finito de puntos (e, incluso, en un conjunto infinito numerable de puntos). Por tanto,

podemos definir la función de densidad como queramos en los puntos en que no exista (que son, a lo sumo, una cantidad finita); ni las probabilidades, ni la función de distribución (que en definitiva es una probabilidad) se modificarán.

Veamos un resultado que fija las condiciones que ha de cumplir una función para que sea de densidad.

Proposición 4.11 *Sea g una función de $\mathbb{R}$ en $\mathbb{R}$ continua salvo, a lo sumo, en un número finito de puntos, tal que*

1. $g(x) \geq 0$ *para todo* $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} g(x)dx = 1$.

Entonces, g es la función de densidad de una variable aleatoria continua.

Demostración. Definamos $G(x) = \int_{-\infty}^{x} g(t)dt$. Es inmediato comprobar que es G es una función de distribución. La teoría de la integración nos dice que G es continua y derivable, siendo su derivada $g(x)$. El resultado queda, pues, demostrado. ♣

Anteriormente habíamos probado que toda función de densidad verificaba estas dos propiedades. Por tanto, dichas propiedades caracterizan a la función de densidad.

Recordando la interpretación geométrica de la integral definida (la integral de una función positiva en un intervalo (a, b) es el área entre dicha función y el segmento (a, b)), podemos *visualizar* algunas de las propiedades y resultados anteriores. Así, la función de densidad es una función positiva (su gráfica está por encima del eje de abscisas) que deja debajo un área igual a la unidad ($\int_{-\infty}^{\infty} f(x)dx = 1$). Es decir:

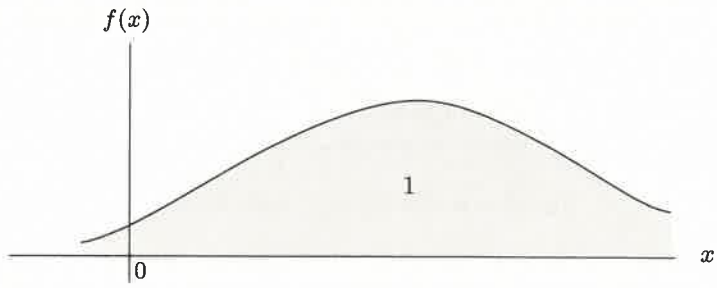


FIG. 4.15.

Asimismo, la función de distribución en x o la probabilidad del intervalo (a, b) será el área que queda *detrás* de x o *encima* de (a, b), respectivamente.

El lector puede entender que *encima* de un punto no hay área, es decir, que $p[X = x]$ es cero.

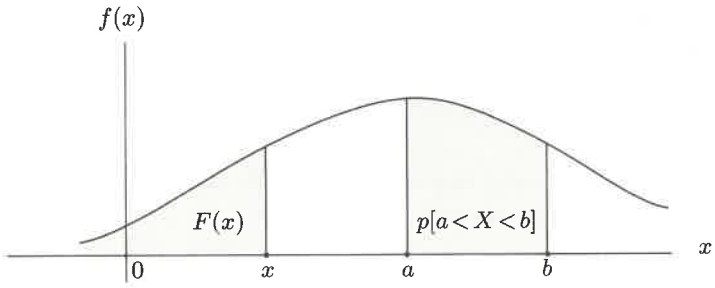


FIG. 4.16.

4.6 Variables aleatorias mixtas

Hasta ahora hemos estudiado las variables aleatorias discretas y las continuas. Las discretas cargan la probabilidad sobre una cantidad finita o infinita numerable de puntos; en las continuas, la probabilidad de cada punto es cero, pero los intervalos aportan probabilidades no nulas. A veces se plantean situaciones en que una variable es continua en un intervalo, pero ciertos puntos tienen probabilidades distintas de cero. Estas situaciones vamos a estudiarlas mediante variables denominadas mixtas, que, como su nombre indica, recogen aspectos modelizados por las variables discretas (algunos puntos aportan cantidades significativas de probabilidad y en ellos la función de distribución *salta, se rompe*) y otros aspectos modelizados por las continuas (tramos de valores con aportaciones continuas de probabilidad, donde la función de distribución crece de forma *suave, sin sobresaltos*). Pasemos a definir este nuevo tipo de variables.

Definición 4.8 *Decimos que una variable aleatoria X es mixta si su función de distribución, F_X, es una combinación lineal convexa de las funciones de distribución correspondientes a una variable aleatoria discreta, F_d, y a una variable aleatoria continua, F_c. Dicho de otra forma, si existe un número α, con $0 \leq \alpha \leq 1$, tal que, para cada número real x, se cumple,*

$$F_X(x) = \alpha \cdot F_d(x) + (1 - \alpha) \cdot F_c(x) \, .$$

Nótese que, para $\alpha = 0$, la expresión anterior corresponde a la función de distribución de una variable aleatoria continua, y, para $\alpha = 1$, a la de una discreta.

Veamos un ejemplo.

Ejemplo 4.10. Supongamos que estamos estudiando la pluviosidad de una zona, y consideramos para ello la variable aleatoria X, *número de litros por m^2 diarios* recogidos en un observatorio. Obviamente, X tomará sólo valores no negativos y, en principio, su comportamiento para los días que llueve podría describirse mediante una función de densidad que fuese nula para los números reales negativos.

Ahora bien, es necesario realizar una pequeña corrección porque el valor $X = 0$ debe tener probabilidad distinta de cero, pues es evidente que el número de días en los cuales no llueve es suficientemente importante.

Propongamos, por ejemplo, la siguiente función de distribución para X:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - 0.5 \cdot e^{-x} & x \geq 0 \end{cases}$$

siendo su gráfica la mostrada en la figura 4.17.

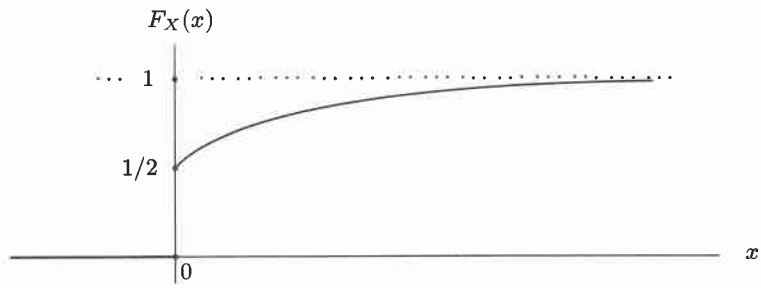


FIG. 4.17.

Advierta el lector que dicha función es, realmente, de distribución, pues verifica las tres propiedades que caracterizan a éstas. Por otra parte, el perfil de esta función difiere de los de las funciones de distribución ya estudiadas. Esta función es una mixtura de las anteriores, presentando un salto en el cero y siendo continua en el resto de los puntos.

También puede reflexionar el lector sobre la conveniencia de una función de este tipo para describir el comportamiento aleatorio de una variable como la estudiada.

Podemos obtener probabilidades a partir de la función de distribución; por ejemplo, la probabilidad de que un día no llueva más de un litro por m^2 será

$$p[X \leq 1] = F_X(1) = 1 - 0.5 \cdot e^{-1} = 0.816 .$$

Por su parte, la probabilidad de que un día llueva valdrá:

$$p[X > 0] = 1 - p[X \leq 0] = 1 - F_X(0) = 0.5 .$$

Obsérvese que para todo número real x se cumple la igualdad

$$F_X(x) = 0.5 \cdot F_d(x) + 0.5 \cdot F_c(x)$$

donde

$$F_d(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

es la función de distribución de la variable discreta que vale siempre la unidad, y

$$F_c(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

es la función de distribución de una variable continua.

Por tanto, la variable X es una variable aleatoria mixta, pues hemos expresado su función de distribución como combinación lineal convexa de las funciones de distribución de una variable discreta y otra continua. ◁

El cálculo de probabilidades para una variable aleatoria mixta se realiza sin excesiva dificultad, al menos para los intervalos; en efecto,

$$p[a < X \leq b] = F_X(b) - F_X(a) = \alpha(F_d(b) - F_d(a)) + (1-\alpha)(F_c(b) - F_c(a)) .$$

Si llamamos f_d a la función de probabilidad asociada a la función de distribución discreta, F_d, y f_c a la función de densidad asociada a la función de distribución continua, F_c, sabemos que

$$F_d(b) - F_d(a) = \sum_{x_i \mid a < x_i \leq b} f_d(x_i) ,$$

y que

$$F_c(b) - F_c(a) = \int_a^b f_c(x)dx .$$

En definitiva, puede obtenerse la probabilidad de que X esté en el intervalo $(a, b]$ sin utilizar las funciones de distribución, siempre que se conozcan la función de probabilidad y la función de densidad correspondientes a las funciones de distribución de la combinación lineal convexa. Así,

$$p[a < X \leq b] = \alpha \sum_{x_i \mid a < x_i \leq b} f_d(x_i) + (1-\alpha) \int_a^b f_c(x)dx .$$

Se puede demostrar la generalidad de este resultado. En concreto, si B es un subconjunto de la σ-álgebra de Borel, entonces,

$$p[X \in B] = \alpha \sum_{x_i \in B} f_d(x_i) + (1-\alpha) \int_B f_c(x)dx .$$

OBSERVACIONES

1) Considerando que una variable mixta es, realmente, discreta en un conjunto, a lo sumo, infinito numerable de puntos y continua en el resto, podríamos haberla definido como *una variable aleatoria X cuya función de distribución F_X es:*

1. *Discontinua por la izquierda en $\mathcal{D} = \{x_1, \ldots, x_n, \ldots\}$, con $x_1 < x_2 < \cdots < x_n < \cdots$.*
2. *En los demás puntos de la recta real, $\mathcal{C} = \mathbb{R} - \mathcal{D}$, es continua y su derivada, no nula, existe y es continua salvo, a lo sumo, en un número finito de puntos.*

Esta definición nos llevaría a introducir el concepto de **función de probabilidad/densidad** de la variable, que recoge la función de probabilidad en la parte discreta (en las discontinuidades de F) y la función de densidad en la continua. Es decir:

$$f_X(x) = \begin{cases} p[X = x] = F_X(x) - F_X(x-) & x \in \mathcal{D} \\ \\ \dfrac{d}{dx} F_X(x) & x \in \mathcal{C} \end{cases}$$

Como se puede intuir, esta función cumple:

1. $f(x) \geq 0$ para todo $x \in \mathbb{R}$,
2. $\sum_{x_i \in \mathcal{D}} p[X = x_i] + \int_{\mathcal{C}} f(x)\, dx = 1$,

propiedades éstas que la caracterizan.

La función de probabilidad/densidad facilita el cálculo de la probabilidad de cualquier suceso, pues

$$p(B) = \sum_{x_i \in B \cap \mathcal{D}} p[X = x_i] + \int_{B \cap \mathcal{C}} f(x)\, dx\ .$$

Como ocurre con las variables aleatorias discretas y continuas, el conocimiento de la función de probabilidad/densidad nos permite obtener la función de distribución correspondiente.

$$F(x) = p[X \leq x] = \sum_{x_i \leq x} p[X = x_i] + \int_{-\infty}^{x} f(x)\, dx\ .$$

Ejemplo 4.11. Calculemos la función de probabilidad/densidad de la variable X, *número de litros por* m^2, del ejemplo anterior. Teniendo en cuenta que la función de distribución de dicha variable sólo es discontinua en el punto cero (único punto donde la variable es *discreta*),

$$f(x) = \begin{cases} 0 & x < 0 \\ p[X = 0] = 0.5 & x = 0 \\ \frac{d}{dx}(1 - 0.5 \cdot e^{-x}) = 0.5 \cdot e^{-x} & x > 0 \end{cases}$$

Adviértase que es imprescindible especificar los puntos en los que la función es de probabilidad y en los que es de densidad. La función

$$g(x) = \begin{cases} 0 & x < 0 \\ 0.5 \cdot e^{-x} & x \geq 0 \end{cases}$$

es la *misma* función que f pero, no siendo función de densidad de una variable continua, adolece de información para ser considerada como la función de probabilidad/densidad de la variable aleatoria X.

Utilizando la función de probabilidad/densidad, podríamos calcular, por ejemplo, la probabilidad de que un día no llueva más de un litro por m^2:

$$p[X \leq 1] = p[X = 0] + \int_0^1 0.5 \cdot e^{-x}\, dx = 0.5 + (0.5 - 0.5 \cdot e^{-1}) = 0.816\ . \triangleleft$$

2) En este capítulo hemos estudiado tres clases de variables aleatorias que, aun no siendo las únicas, sí cubren las necesidades de modelización de la mayor parte de las situaciones. Las variables aleatorias que se salen de esta tipología no sólo son de compleja construcción y difícil desarrollo, sino que, además, son de dudosa justificación práctica.

Una vez planteado este tipo de variables aleatorias, señalaremos, sin embargo, que los desarrollos en capítulos posteriores se realizarán, únicamente, sobre variables aleatorias discretas y continuas. Los resultados obtenidos se harán extensivos a las variables mixtas, teniendo en cuenta su comportamiento como variables discretas en $\mathcal{D}$ y continuas en $\mathcal{C}$.

4.7 Truncamiento de variables aleatorias

Con frecuencia, se desea utilizar una variable aleatoria dentro de un campo de variación más limitado que el que establece su propia definición. Se trataría, por tanto, de construir una nueva variable aleatoria, cuya función de probabilidad, o de densidad, se obtendría a partir de la correspondiente a la variable primitiva: esta nueva función valdría cero fuera del campo de interés y, dentro del mismo, presentaría un perfil semejante al de la función inicial. De hecho, con este procedimiento, la probabilidad del campo de valores que no es de interés se reparte uniformemente en el nuevo intervalo de variación de la variable. La técnica que permite realizar este trabajo se denomina **truncamiento** de la variable aleatoria inicial, y la desarrollaremos mediante un ejemplo.

Ejemplo 4.12. En la modelización de *la renta familiar en millones de pesetas*, podemos utilizar una distribución continua determinada por la siguiente función de densidad:

$$f_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-x} & \text{si } x \geq 0 \end{cases}$$

cuya gráfica está representada en la figura 4.18.

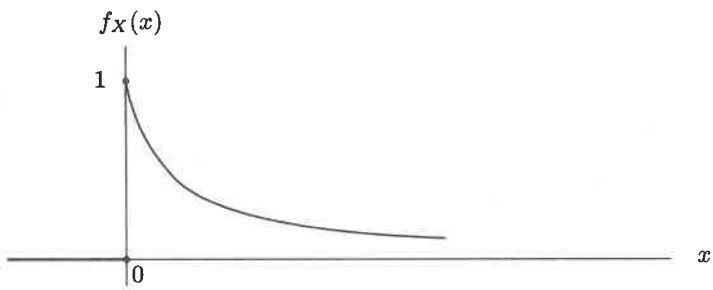


FIG. 4.18.

Nótese que una función de densidad de este tipo recoge la suposición de que un gran porcentaje de familias se concentra en las rentas más bajas, asignando mayores probabilidades a los tramos inferiores de renta (el lector puede comprobar que, por ejemplo, el

porcentaje de familias cuya renta no supera el millón es el 63.21 %); además, la proporción de familias con rentas superiores a una cierta cantidad decrece exponencialmente con la misma.

Para elaborar la Encuesta de Presupuestos Familiares del año 1981, se encuestó a las familias españolas cuya renta se encontraba entre 322575 y 2000000 de pesetas. De acuerdo con la función de densidad propuesta, ese porcentaje representa el 59.1 % de las familias españolas[5].

Podemos estar interesados en estudiar la variable *renta familiar*, X, solamente para este estrato de familias. Llamemos Y a la variable aleatoria *renta de las familias cuya renta se encuentra entre* 0.32 *y* 2 *millones de pesetas.* Como Y toma sus valores en dicho intervalo, su función de distribución cumplirá:

$$F_Y(y) = 0 \text{ si } y < 0.32$$

y

$$F_Y(y) = 1 \text{ si } y > 2 .$$

Dentro del intervalo, esto es, si $0.32 \leq y \leq 2$,

$$F_Y(y) = p[Y \leq y] = p[X \leq y \mid 0.32 \leq X \leq 2] = \frac{p[X \leq y, 0.32 \leq X \leq 2]}{p[0.32 \leq X \leq 2]} =$$
$$= \frac{p[0.32 \leq X \leq y]}{p[0.32 \leq X \leq 2]} = \frac{F_X(y) - F_X(0.32)}{F_X(2) - F_X(0.32)} .$$

La función de distribución de X vale

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x} & \text{si } x \geq 0 \end{cases}$$

como puede comprobarse integrando la función de densidad. Por tanto, $F_X(2) = 0.865$ y $F_X(0.32) = 0.274$, y, sustituyendo en la expresión de la función de distribución de Y, resulta:

$$F_Y(y) = \frac{F_X(y) - 0.274}{0.865 - 0.274} = \frac{F_X(y) - 0.274}{0.591} .$$

De esta forma, hemos obtenido la función de distribución de la nueva variable Y.

$$F_Y(y) = \begin{cases} 0 & y < 0.32 \\ \frac{F_X(y) - 0.274}{0.591} & 0.32 \leq y \leq 2 \\ 1 & y > 2 \end{cases}$$

La función de distribución de Y depende de la de X; sustituyéndola, obtendríamos su expresión explícita.

Al igual que X, la variable aleatoria Y es continua. Podemos obtener su función de densidad; así, en el nuevo intervalo de variación,

$$f_Y(y) = \frac{d}{dy}\left(\frac{F_X(y) - 0.274}{0.591}\right) = \frac{f_X(y)}{0.591} .$$

5. En 1981, el porcentaje real de familias españolas entre estos niveles de renta era el 79 %. La discrepancia es achacable a la inconveniencia del modelo elegido, cuya finalidad es únicamente didáctica.

Entonces,

$$f_Y(y) = \begin{cases} \frac{e^{-y}}{0.591} & 0.32 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

La representación gráfica de esta función de densidad es la representada en la figura 4.19.

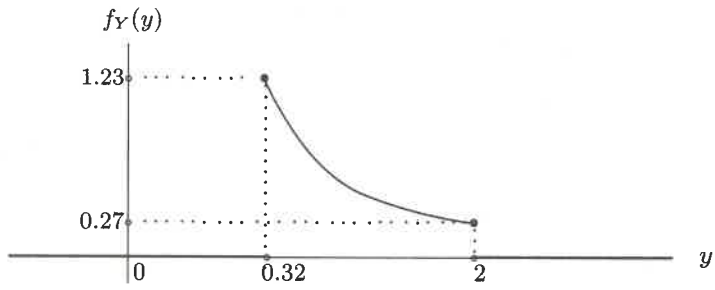


FIG. 4.19.

La comparación de las gráficas correspondientes a las funciones de densidad de X y de Y indica que el reparto uniforme de la probabilidad de los intervalos desechados, 0.409, *eleva* la función de densidad en el intervalo de interés. ◁

El resultado del ejemplo anterior es generalizable a cualquier variable.

Sea X una variable aleatoria e Y la variable resultante del truncamiento de X en el intervalo $(a, b]$, es decir, el truncamiento de X por dos lados. La función de distribución de Y, para los valores de dicho intervalo, es:

$$\begin{aligned} F_Y(y) = p[X \leq y \mid a < X \leq b] &= \frac{p[X \leq y, a < X \leq b]}{p[a < X \leq b]} = \\ &= \frac{p[a < X \leq y]}{p[a < X \leq b]} = \frac{F_X(y) - F_X(a)}{F_X(b) - F_X(a)}, \end{aligned}$$

siendo F_X la función de distribución de X.

Por tanto, teniendo en cuenta que la variable Y no toma valores antes de a y los ha tomado todos después de b, su función de distribución es:

$$F_Y(y) = \begin{cases} 0 & y \leq a \\ \frac{F_X(y) - F_X(a)}{F_X(b) - F_X(a)} & a < y \leq b \\ 1 & y > b \end{cases}$$

Puede comprobar el lector que F_Y es, realmente, una función de distribución.

Si X es continua, lo será también Y. Entonces, se obtiene su función de densidad derivando la de distribución:

$$f_Y(y) = \begin{cases} \frac{f_X(y)}{F_X(b)-F_X(a)} & a < y < b \\ \\ 0 & \text{en el resto} \end{cases}$$

donde f_X es la función de densidad de X.

Asimismo, si X es discreta, Y también lo será, y puede obtenerse directamente su función de probabilidad a partir de la de X. En efecto, si $a < x_i \leq b$,

$$\begin{aligned} p[Y = x_i] &= p[X = x_i \mid a < X \leq b] = \frac{p[X = x_i, a < X \leq b]}{p[a < X \leq b]} = \\ &= \frac{p[X = x_i]}{\sum_{a<x_j\leq b} p[X = x_j]} . \end{aligned}$$

Hemos limitado la variable X al intervalo $(a, b]$. Igualmente, podríamos efectuar el truncamiento de la variable para otro tipo de intervalos: (a, b), $[a, b)$ y $[a, b]$. Con ellos, los resultados anteriores son igualmente válidos, si la variable es continua. En cambio, si es discreta, obtendremos la función de distribución y de probabilidad mediante desarrollos análogos, pero en los que habrá que tener en cuenta la inclusión, o no, de los extremos del intervalo.

En ciertas situaciones, nos interesará sólo acotar, o bien superior, o bien inferiormente, el campo de variación de la variable X; es decir, realizar un truncamiento por un lado de la misma. El intervalo de variación de la variable Y, fruto de dicho truncamiento, será, entonces, $(-\infty, b]$ o (a, ∞). Estas situaciones pueden considerarse casos particulares del truncamiento de X en $(a, b]$, donde $a = -\infty$ o $b = \infty$. Las expresiones obtenidas siguen siendo igualmente válidas, teniendo en cuenta que $F_X(-\infty) = 0$ y que $F_X(+\infty) = 1$. La inclusión, o no, del extremo finito del intervalo de truncamiento, nos llevaría a idénticas consideraciones a las ya realizadas.

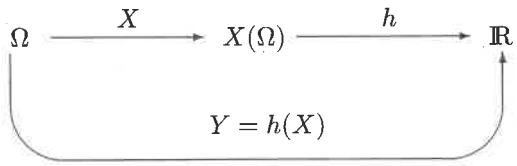
4.8 Transformaciones de una variable aleatoria

En ocasiones, dada la distribución de una cierta variable aleatoria X, estaremos interesados en la distribución de una variable Y, función de X; esto es, realizado un cambio de variable, trataremos de hallar la distribución de la transformada. Así, por ejemplo, conociendo la distribución de X, *número de coches vendidos diariamente* en un concesionario, nos puede interesar la distribución de la variable Y, *beneficios diarios*; obviamente, esta variable vendrá influida por los coches vendidos, X, por el precio unitario del vehículo y por los gastos fijos del establecimiento.

Sea (Ω, α, p) un espacio probabilístico y X una variable aleatoria. Sea h una función cualquiera de $\mathbb{R}$ en sí mismo, es decir, una función real de variable real.

Consideremos la aplicación compuesta $Y = h(X)$, que podemos esquematizar de la siguiente forma:

$$\Omega \xrightarrow{X} \mathbb{R} \qquad \text{y} \qquad \mathbb{R} \xrightarrow{h} \mathbb{R}$$



Es decir, si ω es un punto del espacio muestral Ω,

$$\omega \xrightarrow{X} x = X(\omega) \xrightarrow{h} y = h(x) = h(X(\omega)) .$$

Por tanto, $Y = h(X)$ es una aplicación de Ω en $\mathbb{R}$.

Puede que la función h no esté definida en toda la recta; esto no será relevante siempre que esté definida, al menos, en $X(\Omega)$. Por ejemplo, la aplicación $y = \ln x$ no está definida para los números reales negativos. Por tanto, si X es una variable aleatoria que varía en el intervalo $[-1, 1]$, esto es, si $X(\Omega) = [-1, 1]$, entonces $Y = \ln X$ no estará definida, ya que no existiría $\ln(X(\omega))$ para los puntos $\omega \in \Omega$ tales que $X(\omega) \leq 0$. Ahora bien, si $X(\Omega)$ es el intervalo $[3, 5]$, entonces $3 \leq X(\omega) \leq 5$ para todo punto ω de Ω y, en consecuencia, $\ln(X(\omega))$ está definida siempre.

Aun así, en los desarrollos posteriores consideraremos, por comodidad, h como función de $\mathbb{R}$ en $\mathbb{R}$, si bien debe entenderse que lo determinante es que el campo de definición de la misma incluya $X(\Omega)$.

El primer problema que nos vamos a plantear es si $Y = h(X)$ es, o no, variable aleatoria; en otros términos, ¿cómo debe ser la función real h para que $[h(X) \leq y]$ pertenezca a α para cualquier número real y?

Cuando $Y = h(X)$ sea variable aleatoria, el segundo problema que estudiaremos será su clasificación en discreta, continua o mixta; para ello tendremos en cuenta, tanto qué tipo de variable es X, como qué tipo de función es h.

Por último, trataremos de hallar la distribución de la nueva variable Y, a partir de la distribución conocida de X.

En principio, la mayoría de las funciones de uso habitual, aplicadas sobre una variable aleatoria, proporcionan una nueva variable aleatoria. Concretemos algunos casos en que esto ocurre.

Proposición 4.12 *Sea X una variable aleatoria y h una función real de variable real biyectiva y monótona; entonces, $Y = h(X)$ es variable aleatoria.*

Demostración. Evidentemente, como h es biyectiva y monótona, existe su función inversa h^{-1}, que también es biyectiva y monótona[6].

Si h es creciente, lo será h^{-1}; entonces,

$$[h(X) \leq y] = [X \leq h^{-1}(y)] .$$

Si h es decreciente, igualmente lo será h^{-1} y, por tanto,

$$[h(X) \leq y] = [X \geq h^{-1}(y)] .$$

En ambos casos, $[h(X) \leq y]$ pertenece a α, por ser X una variable aleatoria, lo cual prueba la proposición. ♣

La condición exigida de que h sea biyectiva y monótona es equivalente a la de que sea suprayectiva y estrictamente monótona. Obsérvese que, cuando h es monótona, lo realmente importante para garantizar que exista h^{-1} es que la función h sea inyectiva (esto es, que dos números distintos cualesquiera se apliquen en distintas imágenes), o lo que es lo mismo, que su monotonía sea estricta. Si la aplicación no es suprayectiva, es decir, si $h(\mathbb{R})$ no es todo $\mathbb{R}$, la demostración puede hacerse de forma análoga. Por ejemplo, si $h(\mathbb{R})$ es el intervalo cerrado $[a, b]$ y, por concretar el comentario, suponemos que h es creciente, el subconjunto $[h(X) \leq y]$ de Ω valdrá:

$$[h(X) \leq y] = \begin{cases} \emptyset & \text{si } y < a \\ [X \leq h^{-1}(y)] & \text{si } a \leq y \leq b \\ \Omega & \text{si } y > b \end{cases}$$

siendo, en todo caso, un suceso de α.

La proposición anterior tiene importantes aplicaciones. Por ejemplo, si X es una variable aleatoria lo serán también e^X (obsérvese que esta aplicación no es suprayectiva), X^3, X^5 y, en general, X^n con n impar, y muchas otras funciones de uso frecuente, como $k \cdot X$ y $X + a$, donde k y a son números reales cualesquiera. Asimismo, si X es no negativa, también serán variables aleatorias $\log X$ y $\ln X$.

Cuando h no es monótona, $Y = h(X)$ no es necesariamente variable aleatoria. Con todo, para la mayor parte de las funciones de uso habitual queda garantizado

6. Dada la aplicación $f : M \to N$, en la sección 4.1 (observación **2**) definimos la aplicación imagen recíproca, f^{-1}, de $\mathcal{P}(N)$ en $\mathcal{P}(M)$, por la cual a cada subconjunto B de N se le asocia $f^{-1}(B) = \{a \in M \mid f(a) \in B\}$.

Cuando f es biyectiva, a cada subconjunto de N formado por un solo elemento, $\{b\}$, se le asocia un subconjunto de M formado, también, por un único elemento: $f^{-1}(\{b\}) = \{a\}$, donde $f(a) = b$. Entonces, podemos definir una aplicación que asocia a cada punto de N el punto de M del cual es imagen. Esta aplicación se denomina inversa y se designa con la misma notación que la aplicación imagen recíproca. En definitiva,

$$\begin{array}{rccc} f^{-1}: & N & \to & M \\ & b & \hookrightarrow & a = f^{-1}(b) \end{array}$$

donde $b = f(a)$.

Cuando f es sólo inyectiva, podemos definir de igual forma la aplicación inversa, f^{-1}, si bien en este caso,

$$f^{-1} : f(N) \to M .$$

que lo es, ya que se puede demostrar que, *si X es variable aleatoria y h es continua, $Y = h(X)$ es variable aleatoria*, resultado que no probaremos.

Ejemplo 4.13. Sea X una variable aleatoria cualquiera. Veamos que X^2 también lo es. Obsérvese que $y = x^2$ es una función de $\mathbb{R}$ en $\mathbb{R}$ continua, pero no es ni inyectiva ni suprayectiva ni monótona. Evidentemente,

$$[X^2 \leq y] = \begin{cases} \emptyset & \text{si } y < 0 \\ [-\sqrt{y} \leq X \leq \sqrt{y}] & \text{si } y \geq 0 \end{cases}$$

En el primer caso, $\emptyset \in \alpha$, y, en el segundo, el suceso obtenido pertenecerá a α, ya que puede ponerse como intersección de $[X \leq \sqrt{y}]$ y de $[X \geq -\sqrt{y}]$, sucesos de α.

Por tanto, X^2 es una variable aleatoria. ◁

Estudiadas algunas condiciones que debe cumplir h para que la transformada de una variable aleatoria, $Y = h(X)$, siga siéndolo, nos plantearemos, ahora, obtener su distribución (función de distribución y función de probabilidad, o de densidad) a partir de la de X, según sea ésta discreta o continua.

4.8.1 Cambio de variable para variables discretas

El problema, anteriormente planteado, sobre las condiciones que ha de cumplir h para que $Y = h(X)$ sea una variable aleatoria, se obvia en el caso de que X sea una variable aleatoria discreta.

Proposición 4.13 *Sea X una variable aleatoria discreta. Entonces, $Y = h(X)$ es también una variable aleatoria discreta, sea cual sea la función real de variable real h.*

Demostración. Sea $X(\Omega) = \{x_1, \ldots, x_n, \ldots\}$. Por tanto,

$$[Y \leq y] = \{\omega \in \Omega \mid Y(\omega) = h(X(\omega)) \leq y\} = \bigcup_{x_i \mid h(x_i) \leq y} [X = x_i] \,.$$

Dado que $[X = x_i] \in \alpha$, por ser X una variable aleatoria, el conjunto $[Y \leq y]$, unión de una colección, a lo sumo, infinita numerable de sucesos de α, también pertenecerá a la σ-álgebra.

Además, como h es una aplicación, los valores que toma $Y = h(X)$ han de ser una cantidad, a lo sumo, igual a la de los valores que toma X y, por tanto, una cantidad finita o infinita numerable. Es decir, Y es variable aleatoria y es discreta. ♣

Calculemos las funciones de distribución y de probabilidad de Y a partir de la función de probabilidad de X:

$$F_Y(y) = p[Y \leq y] = \sum_{x_i \mid h(x_i) \leq y} p[X = x_i] = \sum_{x_i \mid h(x_i) \leq y} f_X(x_i)$$

y

$$f_Y(y) = p[Y = y] = \sum_{x_i \mid h(x_i)=y} p[X = x_i] = \sum_{x_i \mid h(x_i)=y} f_X(x_i) .$$

Ejemplo 4.14. Sea X la variable aleatoria que toma los valores -1, 0 y 1 con probabilidades 0.1, 0.3 y 0.6, respectivamente. Hallemos la distribución de la variable $Y = |X|$, aplicación que asocia a cada valor su correspondiente valor positivo. Por tanto, Y toma los valores 0 y 1.

$$f_Y(0) = p[Y = 0] = p[X = 0] = 0.3$$

y

$$f_Y(1) = p[Y = 1] = p[X = -1] + p[X = 1] = 0.7 .$$

Análogamente calcularíamos la función de distribución de Y. ◁

4.8.2 Cambio de variable para variables continuas

Cuando X es una variable aleatoria continua, la situación es diferente. En primer lugar, no toda transformada de X tiene por qué ser variable aleatoria, si bien es cierto que lo será en la mayor parte de los casos. En segundo lugar, cuando $Y = h(X)$ sea variable aleatoria, puede que sea discreta, continua o mixta, dependiendo de la aplicación h.

Si $Y = h(X)$ es una variable aleatoria, formalmente siempre es posible calcular su función de distribución a partir de la función de densidad de X, f_X, aunque este resultado no tiene mucha importancia desde el punto de vista práctico:

$$\begin{aligned} F_Y(y) = p[Y \leq y] &= p[Y = h(X) \in (-\infty, y]] = p[X \in h^{-1}((-\infty, y])] = \\ &= \int_{h^{-1}((-\infty,y])} f_X(x)dx . \end{aligned}$$

En la mayor parte de las situaciones de interés, $Y = h(X)$ es una variable aleatoria continua, existiendo un resultado que, aunque no es general, nos permite, en muchos casos, obtener directamente su función de densidad sin hallar previamente su función de distribución.

Proposición 4.14 *Sea X una variable aleatoria continua cuya función de densidad es $f_X(x)$. Sea $y = h(x)$ una aplicación de $\mathbb{R}$ en $\mathbb{R}$, estrictamente monótona y derivable, tal que su derivada, $h'(x)$, no se anula. Sea Y la transformada de X por h, $Y = h(X)$; entonces, Y es una variable aleatoria continua con función de densidad*

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \left| \frac{dx}{dy} \right| .$$

Demostración. Por las condiciones impuestas a h (estrictamente monótona y continua, al ser derivable), ésta transforma intervalos en intervalos, con lo cual, $h(X(\Omega))$ será toda

la recta real o un intervalo de la misma. Supongamos que $h(X(\Omega))$ es el intervalo $[a, b]$ (la demostración sería análoga si $a = -\infty$ o $b = \infty$).

Por ser h estrictamente monótona, $Y = h(X)$ es variable aleatoria, según vimos en la proposición 4.11 y en la observación posterior.

Supongamos primero que h es creciente, y sea y un valor posible de la función h; entonces,

$$F_Y(y) = p[Y \leq y] = p[h(X) \leq y] = p[X \leq h^{-1}(y)] .$$

En definitiva,

$$F_Y(y) = \begin{cases} 0 & y < a \\ F_X(h^{-1}(y)) & a \leq y \leq b \\ 1 & y > b \end{cases}$$

La variable $Y = h(X)$ es continua, puesto que F_Y es continua y derivable, salvo, a lo sumo, en un número finito de puntos, como compuesta de funciones que lo son (téngase en cuenta que h^{-1} es continua y derivable por serlo h).

Si derivamos F_Y, obtenemos

$$f_Y(y) = \frac{d}{dy} F_X(h^{-1}(y)) = f_X(h^{-1}(y)) \cdot \frac{dx}{dy} \qquad \text{si} \quad a \leq y \leq b ,$$

anulándose fuera del intervalo $[a, b]$. En los extremos del intervalo, puede que la derivada no exista, lo cual, como ya hemos comentado, no será inconveniente; en este caso, asignaremos al punto donde no exista la derivada un valor cualquiera.

Si h es una función decreciente, e y un valor posible de la misma,

$$F_Y(y) = p[Y \leq y] = p[h(X) \leq y] = p[X \geq h^{-1}(y)] .$$

Por tanto,

$$F_Y(y) = \begin{cases} 0 & y > b \\ 1 - F_X(h^{-1}(y)) & a \leq y \leq b \\ 1 & y < a \end{cases}$$

Al igual que en el caso anterior, F_Y es continua y derivable, salvo, a lo sumo, en un número finito de puntos. Entonces,

$$f_Y(y) = \frac{d}{dy}(1 - F_X(h^{-1}(y))) = f_X(h^{-1}(y)) \cdot \left(-\frac{dx}{dy}\right) \qquad \text{si} \quad a \leq y \leq b ,$$

y se anula fuera del intervalo $[a, b]$. Como en el caso anterior, puede que no exista la derivada en los extremos del intervalo.

La función de densidad de $Y = h(X)$ la podemos escribir como

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot \left|\frac{dx}{dy}\right| & \text{si } y \in h(X(\Omega)) \\ 0 & \text{en el resto} \end{cases}$$

expresión que sintetiza ambas situaciones (h creciente y h decreciente). ♣

Aunque la derivada, h', se anule en un número finito de puntos, el resultado sigue siendo igualmente válido. En principio, la función de densidad de Y no existiría en dichos puntos, pudiendo asignársele un valor cualquiera. Por ejemplo, si X toma como valores todos los de la recta e $Y = X^3$, la función h sería $y = x^3$. Aunque h es estrictamente monótona y derivable, su derivada, $h'(x) = 3x^2$, se

anula en $x = 0$. Por tanto, la derivada de $F_Y(y)$ no estaría definida en el punto $y = 0$, con lo que daríamos a f_Y un valor arbitrario en dicho punto.

Ejemplo 4.15. Sea X una variable aleatoria continua e $Y = aX + b$, donde a y b son números reales con $a \neq 0$. Obtengamos la función de densidad de Y en función de la de X.

Si $a > 0$,

$$F_Y(y) = p[aX + b \leq y] = p\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right) .$$

Por tanto,

$$f_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} .$$

Si $a < 0$,

$$F_Y(y) = p[aX + b \leq y] = p\left[X \geq \frac{y-b}{a}\right] = 1 - F_X\left(\frac{y-b}{a}\right) .$$

Entonces,

$$f_Y(y) = \frac{d}{dy}\left[1 - F_X\left(\frac{y-b}{a}\right)\right] = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} .$$

En definitiva, sea cual sea el signo de a,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|} .$$

Advierta el lector que este resultado se obtendría de forma inmediata aplicando la proposición anterior, puesto que la transformación verifica las hipótesis exigidas.

En particular, supongamos que $Y = -2X + 3$ y que

$$f_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Expresando X en función de Y,

$$X = -\frac{Y-3}{2} ,$$

y, teniendo en cuenta que X toma valores entre 0 y 1, deducimos que el campo de la nueva variable Y es el intervalo $[1, 3]$. Entonces,

$$f_Y(y) = -f_X\left(-\frac{y-3}{2}\right) \cdot \left(-\frac{1}{2}\right) = \begin{cases} \frac{1}{2} & 1 \leq y \leq 3 \\ 0 & \text{en el resto} \end{cases}$$

es la función de densidad de la variable transformada Y. ◁

Las condiciones impuestas a la función h en la proposición anterior resultan demasiado restrictivas en muchas ocasiones. Veamos, a continuación, una importante transformación de X que no cumple condiciones tan exigentes.

Ejemplo 4.16. Sea X una variable aleatoria continua e $Y = X^2$. Obtengamos, en primer lugar, la función de distribución de Y a partir de la de X.

$$F_Y(y) = p[X^2 \leq y] = \begin{cases} 0 & y < 0 \\ p[-\sqrt{y} \leq X \leq \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y \geq 0 \end{cases}$$

Obsérvese que, aunque no probásemos explícitamente que $Y = X^2$ es variable aleatoria cuando X lo es (resultado obtenido en el ejemplo 4.13), ello estaría implícito en el cálculo de la función de distribución de Y. En dicho cálculo, hemos utilizado implícitamente el hecho de que la contraimagen por h, esto es, por la función $y = x^2$, de un intervalo $(-\infty, y]$ cualquiera, es igual, o bien al conjunto vacío, o bien al intervalo real $[-\sqrt{y}, \sqrt{y}]$; esta contraimagen es, en cualquier caso, un subconjunto de Borel.

Derivando, se obtiene la función de densidad:

$$f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] \quad \text{si} \quad y \geq 0\,,$$

anulándose cuando $y < 0$.

En particular, si X una variable aleatoria continua cuya función de densidad es:

$$f_X(x) = \frac{1}{2}e^{-|x|} \quad -\infty < x < \infty\,,$$

e $Y = X^2$, entonces,

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}}e^{-\sqrt{y}} & y \geq 0 \end{cases}$$

resultado que se obtiene sin más que sustituir f_X en f_Y. ◁

Por último, existen transformaciones singulares que dan lugar a variables no continuas, partiendo de variables que sí lo son. En el ejemplo siguiente se plantea una de estas situaciones en las que se *discretiza* una variable continua.

Ejemplo 4.17. Sea X una variable aleatoria con función de densidad

$$f_X(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Hallemos la distribución de probabilidad de la variable $Y = [X]$ donde $[\,\cdot\,]$ es la función *parte entera*, que a cada número real le hace corresponder el número entero inmediatamente anterior.

Evidentemente, Y es una variable discreta, que toma como valores todos los números naturales, ya que X sólo toma valores reales positivos. Es decir:

$$X : \left| \begin{array}{ll} 0 & \text{cuando } 0 \leq X < 1 \\ 1 & \text{cuando } 1 \leq X < 2 \\ \vdots & \vdots \\ k & \text{cuando } k \leq X < k+1 \\ \vdots & \vdots \end{array} \right.$$

La función de probabilidad de Y se obtiene mediante un cálculo de probabilidades en la variable X:

$$f_Y(0) = p[Y = 0] = p[0 \leq X < 1] = \int_0^1 e^{-x}dx = -e^{-1} + 1\,,$$

$$f_Y(1) = p[Y = 1] = p[1 \leq X < 2] = \int_1^2 e^{-x}dx = -e^{-2} + e^{-1}$$

y, en general,

$$f_Y(k) = p[Y = k] \quad = \quad p[k \leq X < k+1] = \int_k^{k+1} e^{-x} dx = -e^{-k-1} + e^{-k} .$$

Puede el lector comprobar que f_Y es, realmente, una función de probabilidad. ◁

4.9 Ejercicios propuestos

4.1. En el espacio muestral $\Omega = \{\omega_1, \omega_2, \omega_3\}$ definimos la siguiente aplicación:

$$X(\omega_1) = X(\omega_3) = 1 \quad \text{y} \quad X(\omega_2) = 2 .$$

¿Cuál es la σ-álgebra más pequeña para la cual X es variable aleatoria?

4.2. Se considera el experimento aleatorio consistente en lanzar un dado normal dos veces. Sea X el número de pares obtenidos. Búsquese la σ-álgebra más pequeña en Ω para que X sea una variable aleatoria.

4.3. Sea (Ω, α, p) un espacio probabilístico y A un subconjunto de Ω. Definimos la función indicador de A como la aplicación $I_A : \Omega \to \mathbb{R}$ tal que

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

¿Qué debe exigirse para que I_A sea variable aleatoria?

4.4. Sea un espacio probabilístico (Ω, α, p) y X una aplicación de Ω en $\mathbb{R}$. Demuéstrese:

a) Si X es constante, es una variable aleatoria sea cual sea la σ-álgebra α.
b) Si $\alpha = \{\emptyset, \Omega\}$, X es una variable aleatoria si y sólo si es una aplicación constante.

4.5. Demuéstrese que, para todo $x \in \mathbb{R}$:

a) $(-\infty, x] = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right)$.
b) $(-\infty, x) = \bigcup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n}\right]$.

4.6. Sea (Ω, α, p) un espacio probabilístico y X una aplicación de Ω en $\mathbb{R}$. Pruébese que las siguientes condiciones son equivalentes:

i) $X^{-1}((-\infty, x]) \in \alpha$ para todo x real.
ii) $X^{-1}([a, b]) \in \alpha$ para todo a y b reales con $a < b$.
iii) $X^{-1}((a, b]) \in \alpha$ para todo a y b reales con $a < b$.
iv) $X^{-1}((a, b)) \in \alpha$ para todo a y b reales con $a < b$.
v) $X^{-1}([a, b)) \in \alpha$ para todo a y b reales con $a < b$.

(Sugerencia: expresar los intervalos reales finitos como uniones, o intersecciones, de intervalos con alguno de los extremos no finito y aplicar los resultados de la proposición 4.2.)

4.7. Compruébese que, dada una variable aleatoria X sobre un espacio probabilístico (Ω, α, p), para todo $x \in \mathbb{R}$, el conjunto

$$[X = x] = \{\omega \in \Omega \mid X(\omega) = x\}$$

se puede probabilizar, y hállese esta probabilidad en función de las probabilidades de intervalos de $\mathbb{R}$.

4.8. Dadas las siguientes funciones:

a)

$$F(x) = \begin{cases} \frac{2}{3}e^{x} & \text{si } x \leq 0 \\ 1 - \frac{2}{3}e^{-x} & \text{si } x > 0 \end{cases}$$

b)

$$G(x) = \begin{cases} 0 & \text{si } x < -1 \\ x & \text{si } |x| \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

analícese si son, o no, funciones de distribución.

4.9. Sea el espacio probabilizable (Ω, α), donde Ω es $\mathbb{R}$ y α es la σ-álgebra de Borel, β. Se define una aplicación sobre β a partir de la familia de intervalos $\{(-\infty, x]\}_{x \in \mathbb{R}}$ como

$$p((-\infty, x]) = G(x) ,$$

donde $G(x)$ es la función que cumple las tres propiedades que caracterizan a una función de distribución. Demuéstrese que p es una probabilidad sobre (Ω, α).

4.10. Sea X una variable aleatoria cuya función de probabilidad es

$$f_X(x) = \begin{cases} k \cdot x & x = 1, 2, \ldots, n \\ 0 & \text{en el resto.} \end{cases}$$

a) Hállese el valor de k.

b) Calcúlese la probabilidad de que X tome un valor par.

4.11. Un examen tipo test consta de cinco preguntas con tres posibles opciones cada una. Un alumno contesta al azar las cinco cuestiones. Suponiendo que cada respuesta acertada vale dos puntos, hállese la distribución del número de puntos obtenidos por el alumno.

4.12. Resuélvase el ejercicio anterior en el supuesto de que al alumno se le reste un punto por cada respuesta errónea.

4.13. El número de errores por factura que un contable comete es una variable aleatoria discreta con función de probabilidad

$$p[X = x] = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \quad \text{si} \quad x = 0, 1, \ldots \quad (\lambda > 0) .$$

a) Calcúlese la probabilidad de que no cometa error.
b) ¿Cuál es la probabilidad de que cometa algún error?
c) Sabiendo que ha cometido al menos un error, ¿cuál es la probabilidad de que no cometa más de cinco?

4.14. Sea X la variable aleatoria que designa el número de coches vendidos cada semana en un establecimiento. Se sabe que X tiene la siguiente función de probabilidad:

X	0	1	2	3	4	5	6	7	8 o más
$p[X = x]$	0.04	0.04	k	0.11	0.3	0.23	0.1	0.05	0.03

a) Hállese el valor de k.
b) Determínese la función de distribución de X.
c) Calcúlese $p[2 < X \leq 5]$, $p[X \geq 7]$ y $p[X \leq 6 \mid X > 3]$.

4.15. Un fontanero guarda en su caja de herramientas diez pares de tornillos y tuercas de distinto tamaño, de modo que los elementos de una pareja no son intercambiables con los de otra. Elige dos tornillos y dos tuercas:

a) Hállese la distribución del número de pares acertados.
b) Dibújese la función de distribución y de probabilidad de dicha variable.

4.16. Un ratón de laboratorio tiene dos posibles caminos para salir de un laberinto. Si toma el primero, caminará 15 centímetros hasta encontrar una bifurcación, siendo 30 y 20 centímetros, respectivamente, las distancias que habrá de recorrer si elige una u otra posibilidad. Si por el contrario se decide por el segundo camino, éste le conducirá a la salida, tras recorrer 60 centímetros. Como el investigador ha impregnado la entrada del segundo camino con un atrayente olor a queso, el ratón elegirá esta opción un 85 % de las veces. Hállese la distribución de la variable que expresa la longitud que recorrerá el ratón hasta llegar a la salida.

4.17. Sea g una función de $\mathbb{R}$ en $\mathbb{R}$ que se anula salvo en un conjunto finito, o infinito numerable, $\{x_1, \ldots, x_n, \ldots\}$ y cumple:

i) $0 \leq g(x_i) \leq 1$ para todo x_i.
ii) $\sum_{i=1}^{\infty} g(x_i) = 1$.

Sea $G(x) = \sum_{x_i \leq x} g(x_i)$. Demuéstrese que G es la función de distribución de una variable aleatoria discreta.

4.18. Estúdiese si cada una de las siguientes funciones es, o no, de densidad:

a)

$$f(x) = \begin{cases} 3x(x-1)/2 & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en el resto} \end{cases}$$

b)

$$g(x) = \begin{cases} 1 - |1 - x| & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en el resto} \end{cases}$$

4.19. Sea X una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} 1 - |x| & \text{si } |x| \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Represéntese gráficamente dicha función.

b) Hállese y dibújese la función de distribución.

c) Calcúlense las siguientes probabilidades: $p[X \geq 0]$ y $p[|X| < 1/2]$.

4.20. Dada la función:

$$f(x) = \begin{cases} 1/x & \text{si } 1 \leq x \leq e \\ 0 & \text{en el resto} \end{cases}$$

a) Demuéstrese que es función de densidad.

b) Hállese la función de distribución.

4.21. Sea X una variable aleatoria cuya función de densidad es

$$f_X(x) = \frac{1}{2} \cdot e^{|x|} \qquad -\infty < x < \infty \,.$$

Calcúlese $p[|X| > 2]$.

4.22. Sea X una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ a(1+x) & 0 < x \leq 1 \\ \frac{2}{3} & 1 < x \leq 2 \\ 0 & 2 < x \end{cases}$$

Obténgase:

a) El valor de a para que f_X sea función de densidad.

b) $p[0.5 < X \leq 1.5]$.

4.23. Sea X la variable aleatoria que recoge el número de días que un paciente está inscrito en el primer lugar de una lista de espera de la Seguridad Social para ser operado. La función de densidad de X es

$$f_X(x) = \begin{cases} \frac{1}{30} \cdot e^{-\frac{1}{30}x} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) ¿Cuál será el mínimo número de días que tendrá que esperar para ser intervenido con una probabilidad de 0.95?

b) Si un paciente lleva inscrito 10 días en el primer lugar de la lista, ¿cuál será la probabilidad de que espere, a lo sumo, tres días más?

4.24. Sea g una función de $\mathbb{R}$ en $\mathbb{R}$ continua salvo, a lo sumo, en un número finito de puntos, y tal que

i) $g(x) \geq 0$ para todo $x \in \mathbb{R}$.
ii) $\int_{-\infty}^{\infty} g(x)dx = 1$.

Sea $G(x) = \int_{-\infty}^{x} g(t)dt$. Demuéstrese que G es una función de distribución de una variable aleatoria continua.

4.25. Dada la función:

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2/2 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ (x+1)/4 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

a) Demuéstrese que $F(x)$ es función de distribución.
b) Calcúlense las siguientes probabilidades: $p[X = 1/2]$, $p[X < 1]$, $p[X > 2]$, $p[X = 1]$, $p[1/2 < X < 5/2]$ y $p[X = 1 \mid X < 2]$.

4.26. Sea X una variable aleatoria cuya función de distribución es:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{2}{3} \cdot x & 0 \leq x < \frac{1}{2} \\ \frac{1}{3} \cdot (2x+1) & \frac{1}{2} \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

a) Compruébese que es, realmente, una función de distribución.
b) Hállese la función de probabilidad/densidad.
c) Exprésese F como combinación lineal convexa de una función de distribución discreta y una continua.

4.27. Una máquina fabrica bolas de rodamiento de 5 centímetros de diámetro. Desajustes en la misma hacen que las bolas tengan distintos tamaños. Se ha estimado que los errores en el diámetro siguen una distribución cuya función de densidad es:

$$f(x) = \frac{1}{2} \cdot e^{-|x|} \qquad -\infty < x < \infty \,.$$

Son desechadas todas las bolas que tengan un error superior a ± 0.5 centímetros.

a) Hállese la distribución de probabilidad (función de distribución o de densidad) de los errores en el diámetro de las bolas no rechazadas.
b) ¿Cuál es la probabilidad de que una bola aceptada tenga un diámetro mayor que 5.2 centímetros?

4.28. El número de pacientes, X, que requieren diariamente los servicios de la unidad coronaria en una determinada comarca, se ajusta a la siguiente función de probabilidad:

$$p[X = x] = e^{-5} \cdot \frac{5^x}{x!} \qquad x = 0, 1, 2, \ldots$$

El hospital comarcal solamente tiene capacidad para recibir tres de estos pacientes, desviando el resto a una clínica de la capital.
Hállese la distribución de probabilidad del número de ingresos en la unidad coronaria del hospital comarcal.

4.29. Se ha estimado que el tiempo de vida, en horas, de un cierto componente electrónico sigue una distribución de probabilidad cuya función de densidad es:

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} \cdot e^{-x/8} & x \geq 0 \end{cases}$$

El departamento de control de calidad rechaza todos los componentes que fallen en las primeras tres horas, comercializándose el resto.

a) Hállese la distribución del tiempo de vida de los componentes comercializados.

b) ¿Cuál es la probabilidad de que un componente no rechazado permanezca en funcionamiento más de 12 horas?

4.30. Sea X una variable aleatoria definida sobre un espacio probabilizable (Ω, α). Demuéstrese que $|X|$ es una variable aleatoria.

4.31. Sea X una variable aleatoria cuya función de densidad es:

$$f_X(x) = \begin{cases} x \cdot e^{-x} & x > 0 \\ 0 & \text{en el resto} \end{cases}$$

Hállese la distribución de $Y = e^{-X}$ y $Z = \ln(X + 1)$.

4.32. Obténgase la distribución de $Y = \frac{1}{X}$, donde X es una variable aleatoria cuya función de densidad es:

$$f_X(x) = \begin{cases} |x| & -1 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

4.33. Sea X una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{2} & -1 < x < 0 \\ \frac{e^{-x}}{2} & x \geq 0 \\ 0 & x \leq -1 \end{cases}$$

Hállese la función de densidad de la variable X^2.

4.34. Sea X una variable aleatoria continua cuya función de distribución, F_X, es estrictamente monótona creciente. Obténgase la distribución de la variable $Y = F_X(X)$.

Capítulo 5

VARIABLES ALEATORIAS n-DIMENSIONALES

Como hemos visto en el capítulo anterior, una variable aleatoria es una característica numérica probabilizable de los resultados simples de un experimento aleatorio (número de caras, renta de una familia, beneficios mensuales de una empresa, etc.). Resulta interesante estudiar conjuntamente dos o más características ligadas a un experimento aleatorio, para intentar explicar la posible relación existente entre las mismas, de cara, fundamentalmente, a prever una de ellas a partir de las otras. Así, nos puede interesar estudiar el peso y la altura de una persona, la renta, el consumo y el tamaño de una familia, o la producción y los gastos de mantenimiento de una empresa.

Trataremos de probabilizar dos o más características numéricas de los individuos de una población, es decir, de los resultados de Ω, a partir de la probabilidad definida en dicho espacio muestral. En este proceso seguiremos un esquema similar al de la formalización de las variables aleatorias unidimensionales, por lo cual omitiremos algunos comentarios y observaciones que serán simple extensión de los ya realizados.

Por comodidad, los desarrollos de este capítulo se harán para dos características numéricas (variables aleatorias bidimensionales), generalizándose los resultados al caso n-dimensional.

5.1 Definición de variable aleatoria n-dimensional

Definición 5.1 *Sea (Ω, α, p) un espacio probabilístico y sean X e Y dos variables aleatorias definidas sobre dicho espacio, esto es,*

$$X : \Omega \longrightarrow I\!R \qquad e \qquad Y : \Omega \longrightarrow I\!R \, .$$

Se llama variable aleatoria bidimensional a la aplicación

$$(X, Y) : \Omega \longrightarrow I\!R^2 \, ,$$

que aplica cada punto de Ω, ω, en el punto del plano $(X(\omega), Y(\omega))$.

Obsérvese que una variable aleatoria bidimensional (X,Y) asocia a cada resultado posible del experimento, ω, un punto del plano $\mathbb{R}^2$ que, con la notación habitual de las aplicaciones, se designa como $(X(\omega),Y(\omega))$. La abscisa del punto, $X(\omega)$, es el valor que toma la primera variable, X, y su ordenada, $Y(\omega)$, el valor de la segunda, Y. Es decir, a cada resultado de Ω se le asigna un par de valores: los de las dos características numéricas a estudiar.

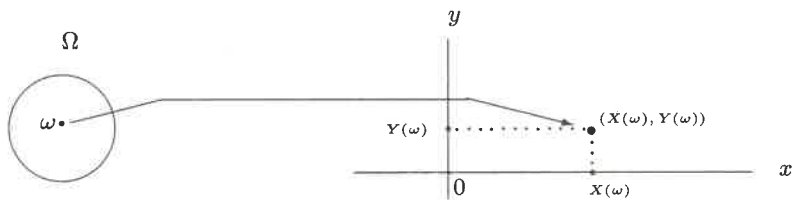


FIG. 5.1.

A veces, interesa considerar a $X(\omega)$ e $Y(\omega)$ como las componentes de un vector, denominándose, en consecuencia, a (X,Y), **vector aleatorio bidimensional**.

Como ocurría para las variables aleatorias unidimensionales, ahora se trata de probabilizar subconjuntos del plano $\mathbb{R}^2$ a partir de las probabilidades dadas en Ω. Hemos de estudiar para qué subconjuntos $B \in \mathbb{R}^2$ se cumplirá que la contraimagen de B por la aplicación (X,Y),

$$(X,Y)^{-1}(B) = \{\omega \in \Omega \mid (X(\omega),Y(\omega)) \in B\} ,$$

es un suceso de Ω, esto es, pertenece a α. La solución de este problema, que no es sencilla, es análoga a la que indicamos para las variables aleatorias unidimensionales, y consiste en que, si B pertenece a una σ-álgebra del plano que denominaremos de Borel, queda garantizado que $(X,Y)^{-1}(B)$ es un suceso y, por tanto, se puede probabilizar. Demos la siguiente definición.

Definición 5.2 *La σ-álgebra de Borel del plano real $\mathbb{R}^2$, que denotaremos por β^2, es la σ-álgebra* más pequeña *que contiene a todos los rectángulos del plano, entendiendo la expresión* más pequeña *en el sentido de que toda σ-álgebra a la que pertenezcan los rectángulos contiene a β^2.*

Los rectángulos del plano son productos cartesianos de intervalos reales. Si consideramos, por ejemplo, los intervalos cerrados $[a,b]$ y $[c,d]$ de la recta, un rectángulo sería

$$[a,b] \times [c,d] = \{(x,y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\} .$$

La figura 5.2 muestra su representación e ilustra por qué se denominan rectángulos a los subconjuntos del plano de este tipo.

Obsérvese que, según cómo sean los intervalos, el rectángulo incluirá todos sus lados, alguno de ellos o ninguno. Por ejemplo, el rectángulo $(a,b) \times (c,d)$ no incluye

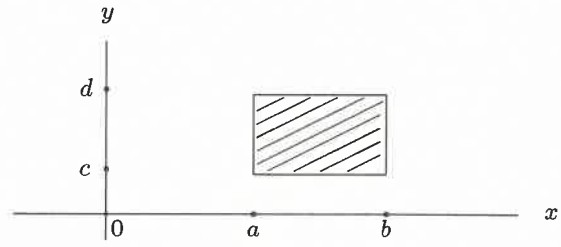


FIG. 5.2.

ninguno, el $(a, b) \times [c, d]$ incluye sólo los lados superior e inferior y el $[a, b] \times [c, d]$, que hemos descrito anteriormente, incluye los cuatro.

Los rectángulos pueden ser ilimitados, si alguno de los intervalos que lo forman lo son. Así,

$$(-\infty, b] \times [c, d] = \{(x, y) \in \mathbb{R}^2 \mid x \leq b, c \leq y \leq d\} .$$

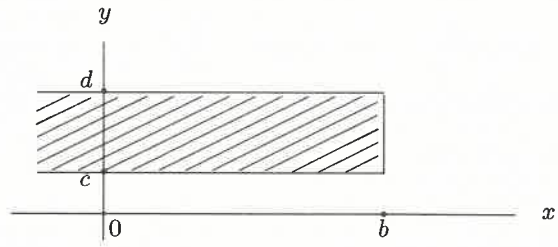


FIG. 5.3.

O bien,

$$[b, \infty) \times [d, \infty) = \{(x, y) \in \mathbb{R}^2 \mid x \geq b, y \geq d\} .$$

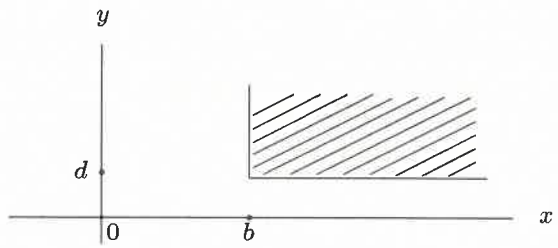


FIG. 5.4.

La σ-álgebra de Borel, β^2, contiene todos los rectángulos del plano. Como es una σ-álgebra, contiene, asimismo, todas las uniones e intersecciones de una colección finita o infinita numerable de rectángulos y los complementarios de ellas. Por tanto, en la σ-álgebra β^2 está contenida cualquier recta o segmento del plano, así como cualquier punto del mismo. Por ejemplo,

$$((-\infty, b] \times (-\infty, \infty)) \cap ([b, \infty) \times (-\infty, \infty))$$

es la recta paralela al eje de ordenadas $x = b$, mientras que

$$((-\infty, b] \times (-\infty, d]) \cap ([b, \infty) \times [d, \infty))$$

es el punto del plano (b, d).

Se puede demostrar que hay subconjuntos del plano que no están en β^2, esto es, que β^2 no coincide con $\mathcal{P}(\mathbb{R}^2)$. Ahora bien, en dicha σ-álgebra están todos los subconjuntos que nos interesa probabilizar, y que son todos los anteriormente descritos.

La importancia de la σ-álgebra de Borel, β^2, viene dada por el hecho de que si $B \in \beta^2$, entonces $(X,Y)^{-1}(B)$ es un suceso, esto es, $(X,Y)^{-1}(B) \in \alpha$. La demostración general desborda los planteamientos de este libro, aunque para ciertos subconjuntos sencillos, y que son precisamente aquellos de nuestro interés, se prueba sin dificultad. Por ejemplo, si $B = [a,b] \times [c,d]$, podemos comprobar que $(X,Y)^{-1}(B)$ pertenece a α. En efecto,

$$\begin{aligned}
(X,Y)^{-1}([a,b] \times [c,d]) &= \{\omega \mid (X(\omega), Y(\omega)) \in [a,b] \times [c,d]\} \\
&= \{\omega \mid X(\omega) \in [a,b], Y(\omega) \in [c,d]\} \\
&= \{\omega \mid X(\omega) \in [a,b]\} \cap \{\omega \mid Y(\omega) \in [c,d]\} \\
&= X^{-1}([a,b]) \cap Y^{-1}([c,d]) \,,
\end{aligned}$$

conjunto que pertenece a α, ya que tanto $X^{-1}([a,b])$ como $Y^{-1}([c,d])$ están en α al ser X e Y variables aleatorias.

Puede probar el lector que, si $B = A \times C$, donde A y C pertenecen a la σ-álgebra de Borel, β, se cumple que

$$(X,Y)^{-1}(B) \in \alpha \,.$$

La demostración es análoga a la anterior, es decir, se comprueba que

$$(X,Y)^{-1}(B) = X^{-1}(A) \cap Y^{-1}(C) \,.$$

Antes de continuar, fijemos la siguiente notación para las contraimágenes de los subconjuntos del plano:

Sea (X,Y) una variable aleatoria bidimensional y B un subconjunto de $\mathbb{R}^2$. Con la notación

$$[(X,Y) \in B]$$

queremos indicar el suceso de Ω

$$(X,Y)^{-1}(B) = \{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in B\} \,.$$

En particular, los sucesos

$$(X,Y)^{-1}([a,b]\times[c,d]) \quad y \quad (X,Y)^{-1}((-\infty,b]\times(-\infty,d]),$$

los escribiremos, respectivamente, como

$$[a\le X\le b, c\le Y\le d] \quad y \quad [X\le b, Y\le d],$$

siguiendo un criterio idéntico para cualquier otro tipo de rectángulo.

Todo lo anterior permite, dada una variable aleatoria bidimensional, trasladar las probabilidades desde Ω al plano real, pudiéndose calcular la probabilidad de que las características en estudio tomen valores en un cierto subconjunto de $\mathbb{R}^2$. Ello se realiza a través de la denominada probabilidad inducida por la variable aleatoria bidimensional que definiremos a continuación.

Definición 5.3 *Sea (Ω,α,p) un espacio probabilístico y (X,Y) una variable aleatoria bidimensional. Definimos la probabilidad inducida por (X,Y) como la aplicación*

$$p_{(X,Y)} : \beta^2 \to \mathbb{R},$$

tal que, para todo B perteneciente a β^2,

$$p_{(X,Y)}(B) = p[(X,Y)^{-1}(B)].$$

Al igual que ocurría en el caso unidimensional, puede comprobarse que la probabilidad inducida por una variable aleatoria bidimensional es, realmente, una probabilidad sobre β^2. La probabilidad inducida, $p_{(X,Y)}$, permite pasar del espacio probabilístico inicial, (Ω,α,p), a un nuevo espacio probabilístico sobre el plano, $(\mathbb{R}^2,\beta^2,p_{(X,Y)})$, que será el adecuado para estudiar la variable aleatoria bidimensional (X,Y).

Ejemplo 5.1. Se lanzan dos dados perfectos de forma independiente y se definen las variables aleatorias X, *resultado del primer dado*, e Y, *menor resultado obtenido entre los dos dados.*

El espacio muestral asociado al experimento está formado por los 36 resultados posibles,

$$\Omega = \{(i,j); i=1,\dots,6, j=1,\dots,6\},$$

cada uno de ellos con probabilidad 1/36.

Las dos variables aleatorias se definen de la siguiente forma:

$$X[(i,j)] = i \quad \text{para todo } (i,j)\in\Omega$$

e

$$Y[(i,j)] = \min(i,j) \quad \text{para todo } (i,j)\in\Omega.$$

Estamos interesados en probabilizar sucesos del tipo

$$\begin{aligned}[X\le 2, Y\le 1] &= (X,Y)^{-1}((-\infty,2]\times(-\infty,1]) = \\ &= \{\omega\in\Omega \mid X(\omega)\in(-\infty,2], Y(\omega)\in(-\infty,1]\} = \\ &= \{\omega\in\Omega \mid X(\omega)\le 2, Y(\omega)\le 1\} = \{(1,1),(2,1)\}.\end{aligned}$$

Ahora, podemos calcular, sin dificultad, probabilidades de sucesos de β^2 a partir de la probabilidad definida sobre Ω. Así, por ejemplo,

$$p_{(X,Y)}((-\infty,2]\times(-\infty,1]) = p[X\leq 2, Y\leq 1] = p[(1,1),(2,1)] = \frac{2}{36}\ .\triangleleft$$

Antes de continuar, generalizaremos las anteriores definiciones para variables n-dimensionales.

Definición 5.4 *Sea (Ω,α,p) un espacio probabilístico y sean $X_1,\ldots,X_n$, n variables aleatorias definidas sobre dicho espacio, esto es,*

$$X_i : \Omega \longrightarrow \mathbb{R}\ , \qquad i=1,\ldots,n\ .$$

Se llama variable aleatoria n-dimensional a la aplicación

$$(X_1,\ldots,X_n) : \Omega \longrightarrow \mathbb{R}^n\ ,$$

que aplica cada punto de Ω, ω, en el punto $(X_1(\omega),\ldots,X_n(\omega))$ de $\mathbb{R}^n$.

La variable aleatoria n-dimensional también recibe el nombre de **vector aleatorio n-dimensional**, denotándose, entonces, como $\mathbf{X}=(X_1,\ldots,X_n)$.

Cuando $n=3$, se utiliza habitualmente la notación (X,Y,Z), en lugar de la propuesta en la anterior definición, (X_1,X_2,X_3).

Se define la *σ-álgebra de Borel en $\mathbb{R}^n$, β^n, como la σ-álgebra* más pequeña *que contiene a los rectángulos n-dimensionales, esto es, a los productos cartesianos de n intervalos reales.* Se puede demostrar que, si B es un subconjunto de $\mathbb{R}^n$ que pertenece a β^n, el subconjunto de Ω

$$(X_1,\ldots,X_n)^{-1}(B) = \{\omega\in\Omega \mid (X_1(\omega),\ldots,X_n(\omega))\in B\}$$

es un suceso, esto es, pertenece a α.

Asimismo, se generaliza la definición de *probabilidad inducida por una variable aleatoria n-dimensional, como la aplicación*

$$p_{(X_1,\ldots,X_n)} : \beta^n \longrightarrow \mathbb{R}\ ,$$

tal que, para todo B perteneciente a β^n,

$$p_{(X_1,\ldots,X_n)}(B) = p[(X_1,\ldots,X_n)\in B]\ ,$$

donde la notación $[(X_1,\ldots,X_n)\in B]$ indica el suceso $(X_1,\ldots,X_n)^{-1}(B)$.

Como puede comprobarse, esta aplicación es, realmente, una probabilidad, y permite construir el espacio probabilístico

$$(\mathbb{R}^n,\beta^n,p_{(X_1,\ldots,X_n)})$$

sobre el que se estudiará el comportamiento aleatorio de una variable n-dimensional, $(X_1,\ldots,X_n)$.

5.2 Función de distribución. Propiedades

Si en el caso unidimensional comentábamos que la probabilidad inducida por una variable aleatoria es de difícil manejo, la afirmación es tanto o más oportuna cuando la variable aleatoria es de dimensión mayor que la unidad. Resulta entonces de interés definir, también ahora, una función numérica, denominada función de distribución, que permita calcular la probabilidad asociada a una variable aleatoria n-dimensional y, en consecuencia, estudiar el comportamiento probabilístico de la misma.

Como en la sección anterior, desarrollaremos primero el caso bidimensional y generalizaremos, sin demostrarlos, los resultados más importantes para variables n-dimensionales. Comencemos definiendo la función de distribución.

Definición 5.5 *Sea (Ω, α, p) un espacio probabilístico y (X, Y) una variable aleatoria bidimensional sobre Ω. Llamamos función de distribución de (X, Y) a la función*

$$F_{(X,Y)} : \mathbb{R}^2 \longrightarrow \mathbb{R} ,$$

donde

$$F_{(X,Y)}(x, y) = p[X \leq x, Y \leq y] ,$$

para todo $(x, y) \in \mathbb{R}^2$.

Obsérvese que la función de distribución está definida para todo punto (x, y) del plano, ya que el conjunto $[X \leq x, Y \leq y]$ está en α, como antes comentamos.

La definición muestra que la función de distribución se puede obtener a partir de la probabilidad inducida por la variable aleatoria bidimensional. En efecto,

$$\begin{aligned} F_{(X,Y)}(x, y) &= p[X \leq x, Y \leq y] = p[(X, Y)^{-1}((-\infty, x] \times (-\infty, y])] = \\ &= p_{(X,Y)}[(-\infty, x] \times (-\infty, y]] . \end{aligned}$$

Por tanto, $F_{(X,Y)}(x, y)$ es la probabilidad inducida por (X, Y) en el rectángulo $(-\infty, x] \times (-\infty, y]$, es decir, la probabilidad inducida *hasta* el punto (x, y).

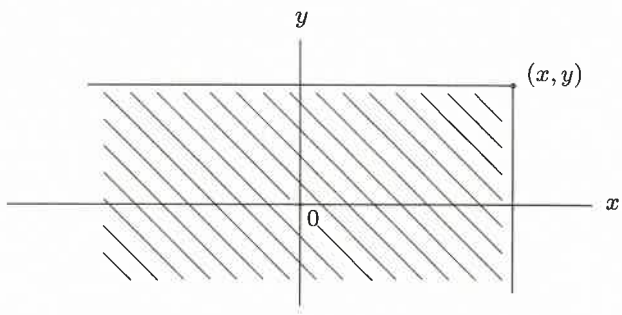


FIG. 5.5.

Cuando no induzca a confusión, suprimiremos el subíndice de la función de distribución para simplificar la notación.

Ejemplo 5.2. Se lanzan al aire dos monedas cargadas, la primera a favor de cara y la segunda a favor de cruz; la probabilidad de *obtener cara* es 0.8 para la primera moneda y 0.4 para la segunda. Denotemos por X a la variable aleatoria *número de caras obtenidas con la primera moneda*, y por Y al *número de caras obtenidas con la segunda.*

En este caso, el espacio muestral es

$$\Omega = \{(c_1, c_2), (c_1, z_2), (z_1, c_2), (z_1, z_2)\} ,$$

donde c_i y z_i designan los resultados *obtener cara y cruz, respectivamente, con la moneda* i. El cálculo de probabilidades en Ω no presenta dificultad. Así, por ejemplo,

$$p[(c_1, c_2)] = p(c_1) \cdot p(c_2) = 0.8 \cdot 0.4 = 0.32 .$$

Puesto que es posible conocer las probabilidades de los sucesos elementales, la σ-álgebra será $\mathcal{P}(\Omega)$. La variable aleatoria bidimensional (X, Y) es la siguiente aplicación de Ω en $\mathbb{R}^2$:

$$\begin{array}{rccc} (X,Y): & \Omega & \to & \mathbb{R}^2 \\ & (c_1, c_2) & \hookrightarrow & (1,1) \\ & (c_1, z_2) & \hookrightarrow & (1,0) \\ & (z_1, c_2) & \hookrightarrow & (0,1) \\ & (z_1, z_2) & \hookrightarrow & (0,0) \end{array}$$

Calculemos su función de distribución para diferentes pares de números reales (x, y). Por ejemplo,

$$\begin{aligned} F_{(X,Y)}(-1,-2) &= p[X \le -1, Y \le -2] = p(\emptyset) = 0 , \\ F_{(X,Y)}(0.5, 0.3) &= p[X \le 0.5, Y \le 0.3] = p[X = 0, Y = 0] = p[(z_1, z_2)] = 0.12 , \\ F_{(X,Y)}(1.5, 0.3) &= p[X \le 1.5, Y \le 0.3] = p[(X = 0, Y = 0) \cup (X = 1, Y = 0)] = \\ &= p[(z_1, z_2)] + p[(c_1, z_2)] = 0.12 + 0.48 = 0.60 , \\ F_{(X,Y)}(0.5, 1.3) &= p[X \le 0.5, Y \le 1.3] = p[(X = 0, Y = 0) \cup (X = 0, Y = 1)] = \\ &= p[(z_1, z_2)] + p[(z_1, c_2)] = 0.12 + 0.08 = 0.20 \quad \text{o} \\ F_{(X,Y)}(2.5, 1.5) &= p[X \le 2.5, Y \le 1.5] = p(\Omega) = 1 . \end{aligned}$$

Formalmente, tendríamos que calcular la función de distribución para cada par de números reales (x, y), del mismo modo que lo hemos hecho para los cinco puntos anteriores; el lector puede comprobar que la función de distribución vale

$$F_{(X,Y)}(x, y) = \begin{cases} 0 & \text{si} \quad x < 0 \text{ o } y < 0 \\ 0.12 & \text{si} \quad 0 \le x < 1, 0 \le y < 1 \\ 0.60 & \text{si} \quad x \ge 1, 0 \le y < 1 \\ 0.20 & \text{si} \quad 0 \le x < 1, y \ge 1 \\ 1 & \text{si} \quad x \ge 1, y \ge 1 \end{cases}$$

Obsérvese que las desigualdades $\le$ y $<$ no están puestas al azar y, por tanto, si cambiamos alguna de ellas, cambia la función.

Esta función de distribución, como cualquier otra correspondiente a una variable aleatoria bidimensional, requiere una representación tridimensional en la que difícilmente se visualiza la distribución de probabilidad en el plano $\mathbb{R}^2$, por lo que normalmente se omite.

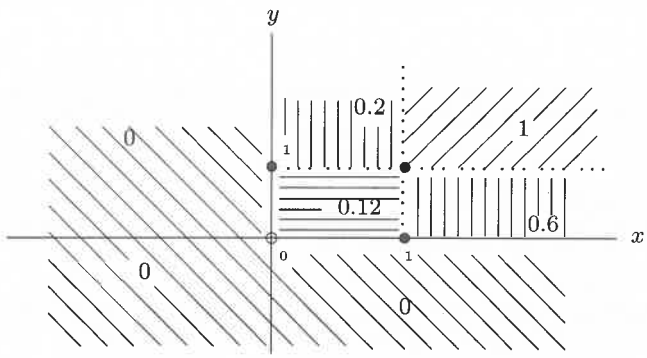


FIG. 5.6.

La figura 5.6 es una representación del plano $\mathbb{R}^2$, especificando el valor de $F_{(X,Y)}$ en las distintas zonas; en cierto sentido, puede entenderse como la gráfica de dicha función a *vista de pájaro*. Esta representación suele tener un interés mayor que el de la propia gráfica tridimensional. ◁

Estudiemos las propiedades de la función de distribución, que enunciamos en la siguiente proposición.

Proposición 5.1 *Sea (X,Y) una variable aleatoria bidimensional y F su función de distribución. Entonces:*

1. $0 \leq F(x,y) \leq 1$ *para todo* $(x,y) \in \mathbb{R}^2$.
2. *F es monótona creciente en cada variable. Es decir,*
   (a) *si $x_1 < x_2$, entonces, $F(x_1,y) \leq F(x_2,y)$ para todo $y \in \mathbb{R}$.*
   (b) *si $y_1 < y_2$, entonces, $F(x,y_1) \leq F(x,y_2)$ para todo $x \in \mathbb{R}$.*
3. $$\lim_{\substack{x\to\infty \\ y\to\infty}} F(x,y) = 1\,.$$
$$\lim_{x\to-\infty} F(x,y) = 0 \quad \textit{para todo} \quad y \in \mathbb{R}\,.$$
$$\lim_{y\to-\infty} F(x,y) = 0 \quad \textit{para todo} \quad x \in \mathbb{R}\,.$$
4. $$\lim_{x\to\infty} F(x,y) = F_Y(y) \quad \textit{para todo} \quad y \in \mathbb{R}\,.$$
$$\lim_{y\to\infty} F(x,y) = F_X(x) \quad \textit{para todo} \quad x \in \mathbb{R}\,.$$

5. *Sean a, b, c y d cuatro números reales tales que $a < b$ y $c < d$. Entonces,*

$$p[a < X \leq b, c < Y \leq d] = F(b,d) - F(a,d) - F(b,c) + F(a,c) .$$

6. *F es continua por la derecha en cada variable. Esto es, para todo $(x,y) \in \mathbb{R}^2$, se cumple*

(a) $\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x+\varepsilon, y) = F(x,y)$.

(b) $\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x, y+\varepsilon) = F(x,y)$.

Demostración. La primera propiedad es evidente, por ser $F(x,y)$ una probabilidad. Probemos la segunda. Si $x_1 < x_2$, entonces

$$[X \leq x_1, Y \leq y] \subset [X \leq x_2, Y \leq y] .$$

Por tanto,

$$p[X \leq x_1, Y \leq y] \leq p[X \leq x_2, Y \leq y]$$

o, lo que es lo mismo, $F(x_1, y) \leq F(x_2, y)$.

El resultado correspondiente a la segunda variable se prueba de manera análoga.

Obsérvese que, aplicando ambas monotonías, se deduce sin dificultad que, si $x_1 < x_2$ e $y_1 < y_2$, se cumple que $F(x_1, y_1) \leq F(x_2, y_2)$.

Para demostrar el tercer resultado[1] obsérvese que, cuando x tiende hacia ∞, $[X \leq x]$ tiende a Ω, y que cuando x tiende hacia $-\infty$, $[X \leq x]$ tiende al suceso $\emptyset$; lo mismo ocurre para Y. En consecuencia, y dado que

$$[X \leq x, Y \leq y] = [X \leq x] \cap [Y \leq y] ,$$

obtenemos

$$\begin{aligned} \lim_{\substack{x \to \infty \\ y \to \infty}} F(x,y) &= \lim_{\substack{x \to \infty \\ y \to \infty}} p([X \leq x] \cap [Y \leq y]) = \\ &= p(\Omega \cap \Omega) = p(\Omega) = 1 . \end{aligned}$$

De manera similar,

$$\begin{aligned} \lim_{x \to -\infty} F(x,y) &= \lim_{x \to -\infty} p([X \leq x] \cap [Y \leq y]) = \\ &= p(\emptyset \cap [Y \leq y]) = p(\emptyset) = 0 . \end{aligned}$$

Cuando y tiende hacia $-\infty$, la demostración es análoga.

Probemos la cuarta propiedad; utilizando los mismos argumentos que en la anterior,

$$\begin{aligned} \lim_{x \to \infty} F(x,y) &= \lim_{x \to \infty} p([X \leq x] \cap [Y \leq y]) = \\ &= p(\Omega \cap [Y \leq y]) = p[Y \leq y] = F_Y(y) \end{aligned}$$

y

$$\begin{aligned} \lim_{y \to \infty} F(x,y) &= \lim_{y \to \infty} p([X \leq x] \cap [Y \leq y]) = \\ &= p([X \leq x] \cap \Omega) = p[X \leq x] = F_X(x) . \end{aligned}$$

1. Las demostraciones de las propiedades *3*, *4* y *6* son heurísticas, al igual, y por los mismos motivos, que lo fueron las correspondientes a la función de distribución de una variable aleatoria unidimensional.

Como puede observarse, esta propiedad permite obtener las funciones de distribución de las variables X e Y a partir de la distribución conjunta de ambas variables. Por ello, tanto la función de distribución de X, como la de Y, se denominan **función de distribución marginal**[2].

Demostremos la quinta propiedad descomponiendo el suceso $[X \leq b, Y \leq d]$ de la siguiente forma:

$$[X \leq b, Y \leq d] = [a < X \leq b, c < Y \leq d] \cup [X \leq a, Y \leq d] \cup [X \leq b, Y \leq c] ,$$

donde el primer suceso de la unión es disjunto de los otros dos y, por tanto, de su unión.

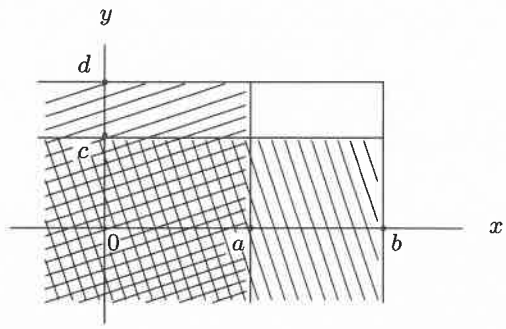


FIG. 5.7.

Tomando probabilidades, obtenemos:

$$p[X \leq b, Y \leq d] = p[a < X \leq b, c < Y \leq d] + p([X \leq a, Y \leq d] \cup [X \leq b, Y \leq c]) .$$

Pero la probabilidad de esa unión vale

$$p[X \leq a, Y \leq d] + p[X \leq b, Y \leq c] - p([X \leq a, Y \leq d] \cap [X \leq b, Y \leq c]) ,$$

es decir,

$$F(a, d) + F(b, c) - F(a, c) ,$$

ya que la intersección de los sucesos $[X \leq a, Y \leq d]$ y $[X \leq b, Y \leq c]$ es el suceso $[X \leq a, Y \leq c]$.

Por tanto,

$$F(b, d) = p[a < X \leq b, c < Y \leq d] + F(a, d) + F(b, c) - F(a, c) ,$$

con lo cual queda demostrada la propiedad *5*.

Esta propiedad nos proporciona la probabilidad de los rectángulos $(a, b] \times (c, d]$, a partir de los valores de la función de distribución en los vértices del mismo.

Cuando el rectángulo no está acotado inferiormente, bien en x, bien en y, el resultado se puede generalizar. En el primer caso, descompongamos el suceso $[X \leq b, Y \leq d]$ como unión de disjuntos:

$$[X \leq b, Y \leq d] = [X \leq b, c < Y \leq d] \cup [X \leq b, Y \leq c]$$

2. Con el fin de simplificar, utilizaremos la notación $F(+\infty, +\infty) = 1$, $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, $F(+\infty, y) = F_Y(y)$ y $F(x, +\infty) = F_X(x)$ para designar los límites a los que se refieren las propiedades *3* y *4* de la proposición 5.1.

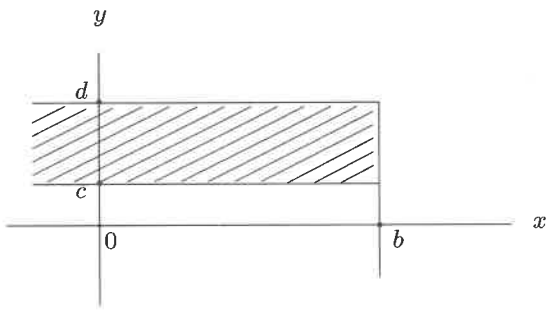


FIG. 5.8.

Tomando probabilidades, resulta

$$p[X \leq b, c < Y \leq d] = F(b,d) - F(b,c) .$$

Igualmente, demostraríamos que

$$p[a < X \leq b, Y \leq d] = F(b,d) - F(a,d) .$$

Por último, probemos la sexta propiedad, en concreto, la continuidad por la derecha para la variable x. Es decir, veamos que

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x+\varepsilon, y) - F(x,y) = 0 .$$

Como hemos demostrado en la propiedad anterior,

$$F(x+\varepsilon, y) - F(x,y) = p[x < X \leq x+\varepsilon, Y \leq y] .$$

Cuando ε tiende a cero, el suceso $[x < X \leq x+\varepsilon]$ tiende a $[x < X \leq x]$, es decir, al suceso $\emptyset$. Por tanto,

$$\begin{aligned} \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} (F(x+\varepsilon, y) - F(x,y)) &= \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} p[x < X \leq x+\varepsilon, Y \leq y] = \\ &= p(\emptyset \cap [Y \leq y]) = p(\emptyset) = 0 . \end{aligned}$$

Con lo cual queda probado que la función de distribución es continua respecto de x. De forma análoga demostraríamos la continuidad por la derecha para la variable y. ♣

OBSERVACIONES

1) Como en el caso unidimensional, la función de distribución describe la acumulación de probabilidad de la variable aleatoria según vamos avanzando en el plano, tanto si lo hacemos hacia la derecha, como si lo hacemos hacia arriba (las palabras *derecha* o *arriba* que acabamos de emplear intentan ser una descripción informal, aunque clara desde el punto de vista intuitivo, de lo que en la anterior proposición se enuncia de manera más formal y precisa). Cuando nos movemos hacia la derecha desde un punto cualquiera, aumentamos el cuadrante cuya probabilidad es la función de distribución, por lo que esta probabilidad aumentará;

otro tanto ocurre si nos movemos hacia arriba. En cualquiera de los dos casos, el crecimiento de la función de distribución refleja la cantidad de la probabilidad de la zona que hemos añadido al cuadrante, por lo que crecimientos fuertes de la función de distribución indicarán que pasamos por una zona de mucha probabilidad y viceversa.

2) La función de distribución no tiene por qué ser continua para cada una de las variables, ya que, bien respecto de x, bien respecto de y, bien respecto de ambas, la función puede ser discontinua por la izquierda en algún punto del plano $\mathbb{R}^2$.

Si ε es un número positivo, por la propiedad *5*, obtenemos:

$$F(x,y) - F(x-\varepsilon, y) = p[x-\varepsilon < X \leq x, Y \leq y] .$$

Ahora bien, cuando ε tiende hacia cero, $[x-\varepsilon < X \leq x]$ tiende al suceso $[X = x]$. Por tanto,

$$\begin{aligned} \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} (F(x,y) - F(x-\varepsilon, y)) &= \lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} p[x-\varepsilon < X \leq x, Y \leq y] = \\ &= p[X = x, Y \leq y] . \end{aligned}$$

Es decir,

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x-\varepsilon, y) = F(x,y) - p[X = x, Y \leq y] .$$

Dado que $p[X = x, Y \leq y]$ no tiene por qué ser cero, la función de distribución puede ser discontinua por la izquierda para la variable x.

Por el mismo razonamiento deduciríamos que la función de distribución puede presentar discontinuidades por la izquierda para la variable y, ya que

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x, y-\varepsilon) = F(x,y) - p[X \leq x, Y = y] .$$

3) Según hemos visto en la propiedad *5*, la función de distribución permite calcular la probabilidad de rectángulos del plano de la forma $(a,b] \times (c,d]$, $(-\infty, b] \times (c,d]$ y $(a,b] \times (-\infty, d]$. Como comprobaremos a continuación, también va a permitir calcular la probabilidad de rectángulos que no estén acotados superiormente en x o en y. Así, por ejemplo, para hallar $p[a < X, c < Y \leq d]$, descompongamos el suceso $[Y \leq d]$ de la siguiente forma:

$$[Y \leq d] = [a < X, c < Y \leq d] \cup [X \leq a, Y \leq d] \cup [Y \leq c] ,$$

donde el primer suceso de la unión es disjunto de los otros dos.

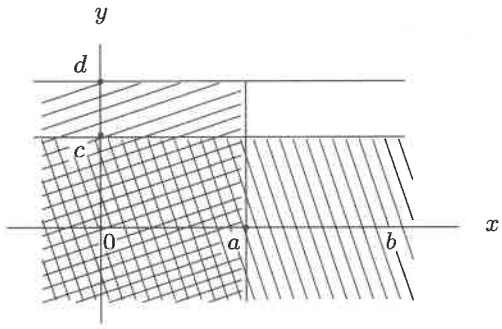


FIG. 5.9.

Entonces,

$$\begin{aligned} p[Y \leq d] &= p[a < X, c < Y \leq d] + p([X \leq a, Y \leq d] \cup [Y \leq c]) = \\ &= p[a < X, c < Y \leq d] + p[X \leq a, Y \leq d] + p[Y \leq c] - p[X \leq a, Y \leq c] , \end{aligned}$$

es decir,

$$p[a < X, c < Y \leq d] = F_Y(d) - F(a, d) - F_Y(c) + F(a, c) .$$

Advierta el lector que, si tuviésemos en cuenta que el límite de una probabilidad es la probabilidad del límite (propiedad cierta, pero no demostrada), el resultado anterior se deduciría de inmediato a partir de la propiedad *5*. Así,

$$\begin{aligned} p[a < X, c < Y \leq d] &= p[a < X < \infty, c < Y \leq d] = \\ &= \lim_{b \to \infty} p[a < X \leq b, c < Y \leq d] = \\ &= \lim_{b \to \infty} (F(b, d) - F(a, d) - F(b, c) + F(a, c)) = \\ &= F(+\infty, d) - F(a, d) - F(+\infty, c) + F(a, c) = \\ &= F_Y(d) - F(a, d) - F_Y(c) + F(a, c) . \end{aligned}$$

Mediante razonamientos análogos hallaríamos la probabilidad de otro tipo de rectángulos. Por ejemplo,

$$\begin{aligned} p[a < X, c < Y] &= p[a < X < \infty, c < Y < \infty] = \\ &= F(+\infty, +\infty) - F(a, +\infty) - F(+\infty, c) + F(a, c) = \\ &= 1 - F_X(a) - F_Y(c) + F(a, c) , \end{aligned}$$

o bien,

$$\begin{aligned} p[X \leq b, c < Y] &= p[-\infty < X \leq b, c < Y < \infty] = \\ &= F(b, +\infty) - F(-\infty, +\infty) - F(b, c) + F(-\infty, c) = \\ &= F_X(b) - F(b, c) . \end{aligned}$$

Cuando estudiamos la función de distribución, F, de una variable aleatoria unidimensional, vimos que había tres propiedades que la caracterizaban:

1. F es monótona creciente.
2. $\lim_{x\to\infty} F(x) = 1$ y $\lim_{x\to-\infty} F(x) = 0$.
3. F es continua por la derecha.

Esto nos puede hacer pensar que las propiedades *2*, *3* y *6* de la proposición anterior caracterizan la función de distribución de una variable aleatoria bidimensional; sin embargo, no es así, como prueba el siguiente ejemplo.

Ejemplo 5.3. Sea F la siguiente función de $\mathbb{R}^2$ en $\mathbb{R}$:

$$F(x,y) = \begin{cases} 1 & x \geq 0, y \geq 0, x+y \geq 1 \\ 0 & \text{en el resto} \end{cases}$$

Puede el lector comprobar que F cumple las propiedades *2*, *3* y *6* de la proposición 5.1, sin más que tener en cuenta cómo es $F(x,y)$ cuando una de las variables, y o x, es constante. Así, $F(x,y)$, como función de x, puede ser:

$$\text{para } y \text{ (fijo) tal que } y < 0, \qquad F(x,y) = 0 \quad -\infty < x < \infty$$

$$\text{para } y \text{ (fijo) tal que } 0 \leq y \leq 1, \qquad F(x,y) = \begin{cases} 0 & -\infty < x < 1-y \\ 1 & 1-y \leq x < \infty \end{cases}$$

$$\text{para } y \text{ (fijo) tal que } y > 1, \qquad F(x,y) = \begin{cases} 0 & -\infty < x < 0 \\ 1 & 0 \leq x < \infty \end{cases}$$

De igual forma describiríamos $F(x,y)$ como función de y.

No obstante, si fuese función de distribución, podríamos calcular la probabilidad de cualquier rectángulo del plano, por ejemplo, de $\left(\frac{1}{3},1\right] \times \left(\frac{1}{3},1\right]$.

La figura 5.10 recoge la disposición de esta función en $\mathbb{R}^2$ y el rectángulo de nuestro interés.

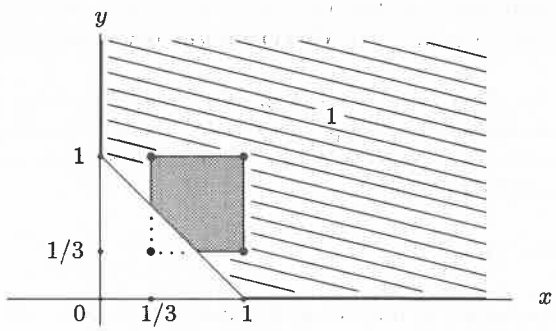


FIG. 5.10.

De acuerdo con la propiedad *5* de dicha proposición, esta probabilidad sería:

$$F(1,1) - F\left(\frac{1}{3},1\right) - F\left(1,\frac{1}{3}\right) + F\left(\frac{1}{3},\frac{1}{3}\right) = 1-1-1+0 = -1\,.$$

Evidentemente, una cantidad negativa no puede ser una probabilidad, es decir, F no verifica la propiedad *5*. En consecuencia, no es una función de distribución de una variable aleatoria bidimensional. ◁

La siguiente proposición, que no demostraremos (el lector puede intuir que esta demostración es paralela a la de la proposición 4.5), nos proporciona las propiedades que caracterizan a la función de distribución de una variable aleatoria bidimensional.

Proposición 5.2 *Sea G una función de $\mathbb{R}^2$ en $\mathbb{R}$. Entonces, G es la función de distribución de una variable aleatoria bidimensional si y sólo si:*

1. G es monótona creciente en cada variable.
2.

$$\lim_{\substack{x\to\infty\\ y\to\infty}} G(x,y) = 1 .$$

$$\lim_{x\to-\infty} G(x,y) = 0 \quad \textit{para todo} \quad y \in \mathbb{R} .$$

$$\lim_{y\to-\infty} G(x,y) = 0 \quad \textit{para todo} \quad x \in \mathbb{R} .$$

3. G es continua por la derecha en cada variable.

4. Para cualquier a, b, c y d, números reales tales que $a < b$ y $c < d$,

$$G(b,d) - G(a,d) - G(b,c) + G(a,c) \geq 0 .$$

Como se comentó para las variables unidimensionales, en la mayoría de las ocasiones determinaremos una variable aleatoria bidimensional asignando directamente una función de distribución (en situaciones especiales veremos que la función puede ser de otro tipo) que describe una cierta disposición de la probabilidad en el plano $\mathbb{R}^2$. Hablaremos, entonces, de **distribución bivariante de probabilidad.**

Ejemplo 5.4. Sea (X,Y) una variable aleatoria bidimensional cuyas componentes, X e Y, miden, respectivamente, *el tiempo, en minutos, que dos cajeras, que llamaremos A y B, tardan en atender a un cliente* en la caja de un supermercado. Asignemos a (X,Y) la siguiente función de distribución:

$$F(x,y) = \begin{cases} (1-e^{-x})\cdot(1-e^{-y}) & \text{si } x \geq 0, y \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

Comprobemos que realmente es función de distribución, es decir, que verifica las cuatro propiedades de la proposición anterior.

Para estudiar la monotonía creciente y la continuidad por la derecha respecto de cada una de las variables, escribamos $F(x,y)$ como función de x (para y constante) y como función de y (para x constante). Así,

$$\text{para } y \text{ (fijo) tal que } y < 0, \qquad F(x,y) = 0 \quad -\infty < x < \infty$$

$$\text{para } y \text{ (fijo) tal que } y \geq 0, \qquad F(x,y) = \begin{cases} (1-e^{-x})\cdot(1-e^{-y}) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Evidentemente, $F(x,y)$ es monótona creciente en x. Cuando $y<0$, es obvio, pues la función es constante. Cuando $y\geq 0$, $F(x,y)$ es constante (en concreto, nula) hasta $x=0$, y creciente, tomando valores positivos, a partir de este valor.

Igualmente, se comprueba la continuidad por la derecha (de hecho, la continuidad, tanto por la derecha, como por la izquierda). Cuando $y<0$, la función es constante y, por tanto, continua. Cuando $y\geq 0$, es continua hasta $x=0$ y a partir de este valor no presenta problemas. En $x=0$, se puede demostrar que

$$\lim_{x\to 0+} F(x,y) = \lim_{x\to 0-} F(x,y) = F(0)\ .$$

Análogamente, se probaría la monotonía creciente y la continuidad respecto de y de la función $F(x,y)$.

Tomando límites, se comprueba sin dificultad que $F(+\infty,+\infty)=1$, $F(x,-\infty)=0$ y $F(-\infty,y)=0$.

Finalmente, si $a<b$ y $c<d$, se cumple que

$$F(b,d)-F(a,d)-F(b,c)+F(a,c)\geq 0\ .$$

Ello es obvio si los cuatro números son iguales a cero. Si los cuatro son no negativos,

$$\begin{aligned}
&F(b,d)-F(a,d)-F(b,c)+F(a,c)= \\
&\quad=(1-e^{-b})\cdot(1-e^{-d})-(1-e^{-a})\cdot(1-e^{-d})- \\
&\qquad-(1-e^{-b})\cdot(1-e^{-c})+(1-e^{-a})\cdot(1-e^{-c})= \\
&\quad=(1-e^{-d})\cdot[(1-e^{-b})-(1-e^{-a})]-(1-e^{-c})\cdot[(1-e^{-b})-(1-e^{-a})]= \\
&\quad=[(1-e^{-d})-(1-e^{-c})]\cdot[(1-e^{-b})-(1-e^{-a})]= \\
&\quad=(e^{-c}-e^{-d})\cdot(e^{-a}-e^{-b})>0\ ,
\end{aligned}$$

ya que, como la función exponencial es creciente, y dado que $a<b$, es decir, $-b<-a$, se verifica que $e^{-b}<e^{-a}$ y, por tanto, $(e^{-a}-e^{-b})>0$. De igual manera se comprueba que $(e^{-c}-e^{-d})$ es también una cantidad positiva.

Si algunos de esos números son iguales a cero y otros no, la comprobación se realizaría de forma análoga.

En definitiva, F cumple las cuatro propiedades de la proposición anterior y, consecuentemente, es una función de distribución. ◁

Los resultados anteriores se generalizan a variables aleatorias n-dimensionales.

Definición 5.6 *Sea (Ω,α,p) un espacio probabilístico y $(X_1,\ldots,X_n)$ una variable aleatoria n-dimensional sobre Ω. Llamamos función de distribución de la variable aleatoria $(X_1,\ldots,X_n)$ a la función*

$$F_{(X_1,\ldots,X_n)}:\mathbb{R}^n\longrightarrow\mathbb{R}\ ,$$

donde

$$F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)=p[X_1\leq x_1,\ldots,X_n\leq x_n]\ ,$$

para todo $(x_1,\ldots,x_n)\in\mathbb{R}^n$.

Cuando no induzca a confusión, suprimiremos el subíndice de $F_{(X_1,\ldots,X_n)}$ para simplificar la notación.

Se comprueba que dicha función de distribución, F, que caracteriza a una **distribución n-variante de probabilidad**, posee propiedades análogas a las demostradas en el caso bidimensional:

1. $0 \leq F(x_1, \ldots, x_n) \leq 1$ para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$.
2. F es monótona creciente en cada variable. Es decir, si $x_i < x_i'$ para $i = 1, \ldots, n$,
$$F(x_1, \ldots, x_i, \ldots, x_n) \leq F(x_1, \ldots, x_i', \ldots, x_n) .$$
3. $F(+\infty, \ldots, +\infty) = 1$ y $F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = 0$ para todo $i = 1, \ldots, n$.
4. Tomando selectivamente límites en la función de distribución conjunta, podemos hallar la función de distribución marginal de cualquier subconjunto de las n variables. Por ejemplo,
$$F(x_1, \ldots, x_h, +\infty, \ldots, +\infty) = F_{(X_1, \ldots, X_h)}(x_1, \ldots, x_h) .$$
En concreto, para todo $i = 1, \ldots, n$,
$$F(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty) = F_{X_i}(x_i) .$$
5. F es continua por la derecha en cada variable. Esto es, para todo $i = 1, \ldots, n$,
$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} F(x_1, \ldots, x_i + \varepsilon, \ldots, x_n) = F(x_1, \ldots, x_i, \ldots, x_n) .$$

Hemos visto cómo la función de distribución permite calcular probabilidades de rectángulos. Se demuestra, además, que es posible obtener la probabilidad de cualquier suceso a partir de ella, aunque el resultado tiene un interés más formal que práctico. Estudiaremos, a continuación, dos tipos de variables para los que existen otros recursos que nos permitirán calcular probabilidades con mayor comodidad.

5.3 Variables n-dimensionales discretas

Definición 5.7 *Sea (Ω, α, p) un espacio probabilístico y X e Y dos variables aleatorias discretas definidas sobre dicho espacio. Decimos entonces que la variable aleatoria (X, Y) es una variable aleatoria bidimensional discreta.*

Como X es una variable aleatoria discreta, el conjunto de valores que toma, $X(\Omega)$, será finito o infinito numerable; otro tanto ocurrirá con $Y(\Omega)$. En los desarrollos de esta sección supondremos que ambos conjuntos son infinito numerables, esto es,
$$X(\Omega) = \{x_1, \ldots, x_n, \ldots\}$$
e
$$Y(\Omega) = \{y_1, \ldots, y_m, \ldots\} ,$$
donde $x_1 < x_2 < \cdots < x_n < \cdots$ e $y_1 < y_2 < \cdots < y_m < \cdots$. Si uno de ellos, o ambos, fueran finitos, el análisis sería similar.

Entonces, los valores que toma la variable aleatoria bidimensional (X, Y) serán, a lo sumo, los del producto cartesiano
$$X(\Omega) \times Y(\Omega) = \{(x_i, y_j); i = 1, \ldots, n, \ldots, j = 1, \ldots, m, \ldots\} .$$

Este conjunto es también infinito numerable como producto cartesiano de dos que lo son.

Obsérvese que la variable bidimensional (X, Y) no tiene por qué tomar necesariamente todos los valores del producto cartesiano anterior. Por ejemplo, no tiene que existir necesariamente un punto ω de Ω tal que $(X(\omega), Y(\omega)) = (x_1, y_1)$. Puesto que, si x_1 es un valor de X, existe un ω de Ω tal que $X(\omega) = x_1$; por el mismo motivo, existe un ω' tal que $Y(\omega') = y_1$. Pero, ω y ω' pudieran ser distintos. En otras palabras, el conjunto de pares de valores que toma la variable bidimensional, $(X, Y)(\Omega)$, está incluido en el conjunto $X(\Omega) \times Y(\Omega)$, pero no tiene por qué coincidir con él. No obstante, simplifica mucho los desarrollos suponer que los valores de la variable aleatoria bidimensional son los del producto cartesiano, $X(\Omega) \times Y(\Omega)$, asignando probabilidades nulas a los puntos de dicho producto que no son valores de la variable.

Para una variable bidimensional discreta, las probabilidades se calculan con comodidad a partir de la denominada función de probabilidad conjunta de X e Y.

Definición 5.8 *Sea (X, Y) una variable aleatoria bidimensional discreta. Su función de probabilidad conjunta es una aplicación,*

$$f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R} ,$$

definida como

$$f_{(X,Y)}(x, y) = p[X = x, Y = y] \qquad \textit{para todo } (x, y) \in \mathbb{R}^2.$$

Observe el lector que la función de probabilidad conjunta, en la que normalmente omitiremos los subíndices, toma los valores

$$f(x_i, y_j) = p[X = x_i, Y = y_j] \qquad \text{para } i = 1, \ldots, n, \ldots , j = 1, \ldots, m, \ldots$$

y es igual a cero para el resto de los puntos del plano.

La figura 5.11 nos muestra la representación gráfica de una función de probabilidad.

Conocida la función de probabilidad, se pueden calcular las probabilidades de los sucesos para una variable aleatoria bidimensional. Demostremos el siguiente resultado, que el lector encontrará análogo al probado para variables aleatorias unidimensionales.

Proposición 5.3 *Sea (X, Y) una variable aleatoria bidimensional y f su función de probabilidad conjunta. Si B es un subconjunto de $\mathbb{R}^2$,*

$$p[(X, Y) \in B] = \sum_{(x_i, y_j) \in B} f(x_i, y_j) .$$

Demostración. Esta igualdad se obtiene inmediatamente sin más que tener en cuenta que

$$[(X, Y) \in B] = \bigcup_{(x_i, y_j) \in B} [X = x_i, Y = y_j] ,$$

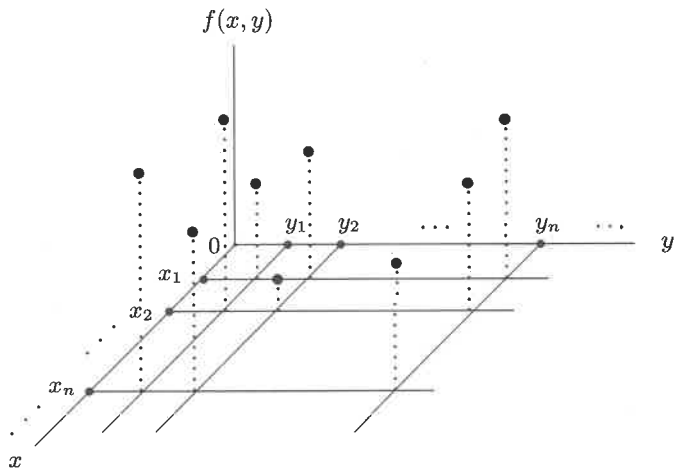


FIG. 5.11.

y que los sucesos de esta unión son disjuntos dos a dos, dado que tanto X como Y son aplicaciones. Puesto que dichos sucesos son una cantidad finita o infinita numerable,

$$p[(X,Y) \in B] = \sum_{(x_i,y_j)\in B} p[X = x_i, Y = y_j] = \sum_{(x_i,y_j)\in B} f(x_i, y_j) \,. \clubsuit$$

La función de distribución, como probabilidad que es, se obtiene a partir de la función de probabilidad teniendo en cuenta la proposición anterior. Asimismo, la función de probabilidad puede hallarse a partir de la función de distribución, según muestra el siguiente resultado.

Proposición 5.4 *Sea (X,Y) una variable aleatoria bidimensional discreta y sean F y f sus funciones de distribución y de probabilidad, respectivamente. Entonces,*

1. $F(x,y) = \sum_{x_i \le x} \sum_{y_j \le y} f(x_i, y_j)$ *para todo* $(x,y) \in \mathbb{R}^2$.
2. $f(x_i, y_j) = F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1})$ *para todo* $i = 2, \ldots, n, \ldots$, $j = 2, \ldots, m, \ldots$

Demostración. En efecto, como

$$[X \le x, Y \le y] = \bigcup_{x_i \le x} \bigcup_{y_j \le y} [X = x_i, Y = y_j] \,,$$

dicho suceso es unión de una colección finita o infinita numerable de sucesos disjuntos dos a dos. Por tanto,

$$F(x,y) = \sum_{x_i \le x} \sum_{y_j \le y} p[X = x_i, Y = y_j] = \sum_{x_i \le x} \sum_{y_j \le y} f(x_i, y_j) \,.$$

Recíprocamente,

$$f(x_i, y_j) = p[X = x_i, Y = y_j] = p[x_{i-1} < X \leq x_i, y_{j-1} < Y \leq y_j] ,$$

y basta aplicar la propiedad *5* de la proposición 5.1 para obtener el resultado. ♣

Obviamente, el segundo resultado de la anterior proposición sigue siendo válido cuando $i = 1$ o $j = 1$, sin más que tener en cuenta las propiedades de la función de distribución cuando una de las variables tiende a $-\infty$. Así, por ejemplo,

$$\begin{aligned} f(x_1, y_j) &= p[X = x_1, Y = y_j] = p[-\infty < X \leq x_1, y_{j-1} < Y \leq y_j] = \\ &= F(x_1, y_j) - F(x_1, y_{j-1}) , \end{aligned}$$

para todo $j = 2, \ldots, m, \ldots$ Igualmente obtendríamos:

$$f(x_i, y_1) = F(x_i, y_1) - F(x_{i-1}, y_1) , \quad \text{para todo} \quad i = 2, \ldots, n, \ldots$$

y

$$f(x_1, y_1) = F(x_1, y_1) .$$

Podría interesarnos conocer el aspecto que presenta la función de distribución de una variable aleatoria bidimensional discreta e, incluso, su gráfica. Ahora bien, como función de $\mathbb{R}^2$ en $\mathbb{R}$, exigiría una representación tridimensional simulada, con los inconvenientes que esto acarrea. Por otro lado, esta posible visualización no aporta una idea inmediata de la distribución de probabilidad. Aun así, indicaremos que F es una función escalonada, siendo los escalones superficies rectangulares dispuestas a distintos niveles, y crecientes tanto en x como en y.

Puesto que la función de probabilidad permite obtener la probabilidad de cualquier suceso en una variable bidimensional discreta, con frecuencia se asigna directamente a este tipo de variables una función de probabilidad que explique su comportamiento aleatorio. Resulta importante, pues, estudiar cuál es el aspecto característico de la función de probabilidad.

Proposición 5.5

1. *Sea f la función de probabilidad de una variable bidimensional discreta, (X, Y), que toma los valores (x_i, y_j), para $i = 1, \ldots, n, \ldots$, $j = 1, \ldots, m, \ldots$ Entonces,*

   (a) $0 \leq f(x, y) \leq 1$ *para todo* $(x, y) \in \mathbb{R}^2$.

   (b) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f(x_i, y_j) = 1$.

2. *Además, si $g : \mathbb{R}^2 \to \mathbb{R}$ es una función que se anula salvo para un conjunto finito o infinito numerable, $\{(x_i, y_j); i = 1, \ldots, n, \ldots, j = 1, \ldots, m, \ldots\}$, y cumple*

   (a) $0 \leq g(x_i, y_j) \leq 1$ *para todo* (x_i, y_j),

   (b) $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) = 1$,

entonces, g es la función de probabilidad de una variable aleatoria bidimensional discreta.

Demostración. Probemos el apartado *1*. La primera propiedad es inmediata ya que los valores de f son probabilidades.

Para probar la segunda, obsérvese que

$$\Omega = \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{\infty}[X = x_i, Y = y_j] ,$$

y, por tanto, Ω es unión de una colección infinita numerable de sucesos disjuntos dos a dos. Consecuentemente,

$$\begin{aligned}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f(x_i, y_j) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p[X = x_i, Y = y_j] = \\ &= p\left[\bigcup_{i=1}^{\infty}\bigcup_{j=1}^{\infty}[X = x_i, Y = y_j]\right] = p(\Omega) = 1 .\end{aligned}$$

Omitiremos la demostración del apartado *2* (el esquema de la misma es similar al del apartado *2* de la proposición 4.8). Por tanto, estas dos propiedades caracterizan a una función de probabilidad, esto es, una función que las cumpla lo es. ♣

Recordemos que, para una variable aleatoria bidimensional cualquiera, (X, Y), podía obtenerse la función de distribución marginal de cada una de las variables, X e Y, tomando límites adecuadamente. Si (X, Y) es discreta, y se conoce su función de probabilidad, pueden obtenerse, asimismo, las funciones de probabilidad marginales de ambas variables, que serán también discretas, sumando valores de la función de probabilidad conjunta. Veamos este resultado.

Proposición 5.6 *Sea (X, Y) una variable aleatoria bidimensional discreta y f su función de probabilidad conjunta. Entonces,*

$$f_X(x_i) = \sum_{j=1}^{\infty} f(x_i, y_j)$$

y

$$f_Y(y_j) = \sum_{i=1}^{\infty} f(x_i, y_j) .$$

Demostración. Basta tener en cuenta que

$$[X = x_i] = \bigcup_{j=1}^{\infty}[X = x_i, Y = y_j]$$

e

$$[Y = y_j] = \bigcup_{i=1}^{\infty}[X = x_i, Y = y_j] ,$$

y que ambas son uniones de una colección finita, o infinita numerable, de sucesos disjuntos dos a dos. Tomando probabilidades, se concluye el resultado. ♣

Los cálculos de probabilidad en variables aleatorias bidimensionales, especialmente cuando las dos componentes, X e Y, toman un número finito de valores, se pueden realizar con comodidad trabajando de forma análoga a como se hacía con las variables estadísticas, es decir, a partir de una tabla de doble entrada en cuyas casillas aparecen los valores de la función de probabilidad en lugar de las frecuencias. Así,

$X \setminus Y$	y_1	$\cdots$	y_j	$\cdots$	y_m	$\cdots$	
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	$\cdots$	$p_{1\bullet}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{im}	$\cdots$	$p_{i\bullet}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
x_n	p_{n1}	$\cdots$	p_{nj}	$\cdots$	p_{nm}	$\cdots$	$p_{n\bullet}$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
	$p_{\bullet 1}$	$\cdots$	$p_{\bullet j}$	$\cdots$	$p_{\bullet m}$	$\cdots$	1

donde, para $i = 1, \ldots, n, \ldots$, $j = 1, \ldots, m, \ldots$,

$$p_{ij} = p[X = x_i, Y = y_j] = f(x_i, y_j) \,,$$

$$p_{i\cdot} = p[X = x_i] = f_X(x_i) = \sum_{j=1}^{\infty} f(x_i, y_j) = \sum_{j=1}^{\infty} p_{ij}$$

y

$$p_{\cdot j} = p[Y = y_j] = f_Y(y_j) = \sum_{i=1}^{\infty} f(x_i, y_j) = \sum_{i=1}^{\infty} p_{ij} \,,$$

obteniéndose los valores de las funciones de probabilidad marginales de X y de Y sin más que sumar filas y columnas, respectivamente.

El siguiente ejemplo, con X e Y tomando ambas un número finito de valores, ilustra esta forma de trabajar.

Ejemplo 5.5. Realizamos el experimento consistente en lanzar dos dados perfectos y consideramos la variable aleatoria bidimensional (X, Y), donde X es el *resultado del primer dado*, e Y el *menor resultado obtenido entre los dos dados* (esta variable aleatoria ha sido propuesta en el ejemplo 5.1). Evidentemente, (X, Y) es una variable discreta, pues X e Y lo son. Puesto que tanto X como Y toman los valores $1, 2, 3, 4, 5$ y 6, la variable (X, Y) tomará, en principio, los pares de valores (i, j), donde $i = 1, \ldots, 6$, $j = 1, \ldots, 6$.

La función de probabilidad conjunta se puede calcular a partir de la probabilidad definida sobre Ω. Así,

$$\begin{aligned} p[X = 1, Y = 1] &= p[(1,1), (1,2), \ldots, (1,6)] = \frac{6}{36} \,, \\ p[X = 1, Y = 2] &= p(\emptyset) = 0 \quad \text{o} \,, \end{aligned}$$

$$p[X=2,Y=1] \quad = \quad p[(2,1)] = \frac{1}{36} \,.$$

De la misma forma, calcularíamos la función de probabilidad en el resto de los valores posibles de (X,Y). Todo ello suele presentarse en una tabla de doble entrada, donde aparecen los valores de dicha función. En nuestro ejemplo:

$X \setminus Y$	1	2	3	4	5	6	
1	6/36	0	0	0	0	0	1/6
2	1/36	5/36	0	0	0	0	1/6
3	1/36	1/36	4/36	0	0	0	1/6
4	1/36	1/36	1/36	3/36	0	0	1/6
5	1/36	1/36	1/36	1/36	2/36	0	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	11/36	9/36	7/36	5/36	3/36	1/36	1

Se puede comprobar que esta tabla proporciona realmente una función de probabilidad conjunta; así, todos los valores que aparecen dentro de la misma están comprendidos entre 0 y 1 y su suma es la unidad.

En la última columna y en la última fila de la tabla, se encuentran los valores de la función de probabilidad marginal de X y de Y, respectivamente, obtenidos al sumar las correspondientes filas y columnas.

Determinada la función de probabilidad, pueden hallarse probabilidades relativas a (X,Y). Por ejemplo,

$$\begin{aligned} p[2.5 < X \leq 4, \;\; 1 \leq Y < 3] \quad &= p[X=3,Y=1] + p[X=4,Y=1] + \\ &\quad + p[X=3,Y=2] + p[X=4,Y=2] = \tfrac{4}{36} \,. \end{aligned}$$

Del mismo modo, el lector podría plantearse obtener la función de distribución calculando $p[X \leq x, Y \leq y]$ para todos los puntos (x,y) del plano, si bien su escaso interés práctico y su complicado cálculo no hacen aconsejable el empeño. ◁

Por último, planteemos en el siguiente ejemplo una variable bidimensional discreta en la que, tanto X como Y, toman un número infinito numerable de valores.

Ejemplo 5.6. Sea (X,Y) una variable aleatoria bidimensional discreta cuya función de probabilidad conjunta es:

$$f(x,y) = \begin{cases} \frac{1}{12} \cdot \left(\frac{2}{3}\right)^{x-1} \cdot \left(\frac{3}{4}\right)^{y-1} & \text{si } x = 1,2,\ldots,\; y = 1,2,\ldots \\ 0 & \text{en el resto} \end{cases}$$

Comprobemos que se trata realmente de una función de probabilidad. Por un lado, $0 \leq f(x,y) < 1$, pues, o bien toma el valor cero, o bien es el producto de tres números comprendidos entre cero y uno. Por otro lado,

$$\begin{aligned} \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} f(x,y) \quad &= \quad \sum_{x=1}^{\infty}\sum_{y=1}^{\infty} \frac{1}{12} \cdot \left(\frac{2}{3}\right)^{x-1} \cdot \left(\frac{3}{4}\right)^{y-1} = \\ &= \quad \frac{1}{12} \cdot \sum_{x=1}^{\infty} \left(\frac{2}{3}\right)^{x-1} \cdot \sum_{y=1}^{\infty} \left(\frac{3}{4}\right)^{y-1} = \frac{1}{12} \cdot 3 \cdot 4 = 1 \,. \end{aligned}$$

Calculemos las funciones de probabilidad marginales. Así,

$$f_X(x) = \sum_{y=1}^{\infty} f(x,y) = \sum_{y=1}^{\infty} \frac{1}{12} \cdot \left(\frac{2}{3}\right)^{x-1} \cdot \left(\frac{3}{4}\right)^{y-1} = \frac{1}{3} \cdot \left(\frac{2}{3}\right)^{x-1} \quad \text{si } x = 1, 2, \dots ,$$

valiendo cero en el resto. De igual forma,

$$f_Y(y) = \frac{1}{4} \cdot \left(\frac{3}{4}\right)^{y-1} \quad \text{si } y = 1, 2, \dots ,$$

y anulándose en el resto.

Conocida la función de probabilidad conjunta, podemos calcular la probabilidad de cualquier suceso expresado en términos de (X, Y). Por ejemplo,

$$p[3 \leq X \leq 5, 15 \leq Y \leq 20] = \sum_{x=3}^{5} \sum_{y=15}^{20} \frac{1}{12} \cdot \left(\frac{2}{3}\right)^{x-1} \cdot \left(\frac{3}{4}\right)^{y-1} = 0.00458 .$$

El lector puede realizar todos los cálculos anteriores utilizando una tabla de doble entrada. ◁

Los resultados obtenidos se extienden, sin dificultad, a variables discretas n-dimensionales.

Definición 5.9 *Sea (Ω, α, p) un espacio probabilístico y $X_1, \dots, X_n$, n variables aleatorias discretas definidas sobre dicho espacio. Decimos, entonces, que la variable aleatoria $(X_1, \dots, X_n)$ es una variable aleatoria n-dimensional discreta.*

Las variables n-dimensionales discretas se manejan con sencillez utilizando *la función de probabilidad conjunta que se define como una función,*

$$f_{(X_1,\dots,X_n)} : \mathbb{R}^n \to \mathbb{R} ,$$

tal que

$$f(x_1, \dots, x_n) = p[X_1 = x_1, \dots, X_n = x_n] \quad \textit{para todo } (x_1, \dots, x_n) \in \mathbb{R}^n .$$

Esta función, en la que omitiremos normalmente los subíndices, cumple propiedades análogas a las obtenidas en el caso bidimensional, que aquí no demostraremos y que enunciamos a continuación:

1. $F(x_1^0, \dots, x_n^0) = \sum_{x_1 \leq x_1^0} \cdots \sum_{x_n \leq x_n^0} f(x_1, \dots, x_n)$.
2. Si B es un subconjunto de $\mathbb{R}^n$,
$$p[(X_1, \dots, X_n) \in B] = \sum_{(x_1,\dots,x_n) \in B} f(x_1, \dots, x_n) .$$
3. $0 \leq f(x_1, \dots, x_n) \leq 1 \quad$ para todo $\quad (x_1, \dots, x_n) \in \mathbb{R}^n$.
4. $\sum_{x_1} \cdots \sum_{x_n} f(x_1, \dots, x_n) = 1$.
5. $f_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} f(x_1, \dots, x_n)$ para todo $i = 1, \dots, n$.

Las propiedades tercera y cuarta caracterizan la función de probabilidad de una variable aleatoria n-dimensional discreta.

5.4 Variables n-dimensionales continuas

Definición 5.10 *Sea (X,Y) una variable aleatoria bidimensional y F su función de distribución. Decimos que (X,Y) es una variable aleatoria bidimensional continua si*

1. *F es una función continua en cada variable.*
2. *Existen sus primeras derivadas parciales, $\frac{\partial}{\partial x}F(x,y)$ y $\frac{\partial}{\partial y}F(x,y)$. Asimismo, su segunda derivada parcial respecto de ambas variables, $\frac{\partial^2}{\partial x \partial y}F(x,y)$, existe y es continua salvo, tal vez, a lo largo de un número finito de curvas.*
3. $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial^2}{\partial x\partial y}F(x,y)dxdy = 1\ .$$

Las variables bidimensionales continuas toman valores en todo el plano real o en un subconjunto del mismo. Pequeñas variaciones en x o en y producen pequeñas variaciones en $F(x,y)$ (de hecho, el rectángulo $[X \leq x, Y \leq y]$, cuya probabilidad es la función de distribución, varía al variar x o y). Es por ello que ninguno de los puntos del plano $\mathbb{R}^2$ aporta cantidades significativas de probabilidad. La continuidad de $F(x,y)$ es la propiedad relevante de las variables continuas, dando nombre a las mismas; las otras dos condiciones de la definición, imprescindibles como veremos posteriormente, se exigen para hacer operativas a este tipo de variables.

Como se ha comentado anteriormente, la función de distribución es de difícil manejo en el cálculo de las probabilidades. De ahí que sea conveniente definir una nueva función que nos permita trabajar con comodidad.

Definición 5.11 *Sea (X,Y) una variable aleatoria bidimensional continua. Su función de densidad conjunta es una aplicación*

$$f : \mathbb{R}^2 \to \mathbb{R}\ ,$$

definida como

$$f(x,y) = \frac{\partial^2}{\partial x\partial y}F(x,y)\ ,$$

donde F es la función de distribución de (X,Y).

La función de densidad conjunta, que cuando se preste a confusión denotaremos por $f_{(X,Y)}$, es una función de $\mathbb{R}^2$ en $\mathbb{R}$ continua, que tal vez no exista, o no sea

continua, en un número finito de curvas[3]. Más adelante veremos que las probabilidades para una variable aleatoria bidimensional continua se obtienen integrando la función de densidad conjunta. Por ello, su indefinición en ciertas curvas no nos creará problemas desde el punto de vista probabilístico, ya que podemos asignar valores arbitrarios a la función de densidad en las mismas.

Se puede comprobar que

$$\frac{\partial^2}{\partial x \partial y} F(x,y) = \frac{\partial^2}{\partial y \partial x} F(x,y) \ ,$$

es decir, que es indiferente el orden de derivación de la función de distribución. Asimismo, se demuestra que es indiferente el orden de integración de la función de densidad cuando se integre en subconjuntos del plano.

El hecho de que la función de densidad conjunta no exista en alguna curva ocurrirá a veces para variables aleatorias que no tomen como valores todos los del plano. En las curvas que separan los valores que toma la variable de los que no toma, es posible que la función de densidad no exista. Veamos el siguiente ejemplo.

Ejemplo 5.7. Se eligen al azar y separadamente dos números en el intervalo $(0,1)$. Sean X e Y las variables aleatorias que indican cada una de las elecciones y consideremos la variable aleatoria bidimensional (X,Y).

Los valores que toma dicha variable son los del rectángulo $(0,1) \times (0,1)$ de $\mathbb{R}^2$, que se denomina, por su aspecto, el *cuadrado unidad*. Puesto que los puntos se escogen al azar, no parece razonable primar a unas zonas del cuadrado unidad sobre otras, por lo que resulta conveniente asignar a (X,Y) la siguiente función de distribución:

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ o } y < 0 \\ x \cdot y & \text{si } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ x & \text{si } 0 \leq x \leq 1,\ y > 1 \\ y & \text{si } x > 1,\ 0 \leq y \leq 1 \\ 1 & \text{si } x > 1,\ y > 1 \end{cases}$$

Aunque posteriormente justificaremos con argumentos más formales la elección de esta función de distribución, obsérvese que asigna como probabilidad de un subconjunto del cuadrado unidad su área y que, por tanto, no prima unas zonas sobre otras. Por ejemplo, si $\mathcal{R}$ es el rectángulo $(0.2, 0.3] \times (0.6, 0.8]$,

$$\begin{aligned} p[(X,Y) \in \mathcal{R}] &= p[0.2 < X \leq 0.3, 0.6 < Y \leq 0.8] = \\ &= F(0.3,0.8) - F(0.2,0.8) - F(0.3,0.6) + F(0.2,0.6) = \\ &= (0.3)\cdot(0.8) - (0.2)\cdot(0.8) - (0.3)\cdot(0.6) + (0.2)\cdot(0.6) = 0.02 \ , \end{aligned}$$

que, según puede comprobarse, es el área del mismo.

3. Formalmente habría que precisar que las curvas en las que la función de densidad conjunta tal vez no exista, o no sea continua, no pueden ser cualesquiera, sino que su gráfica en el plano tiene que tener *contenido nulo*. Este concepto no va a ser desarrollado en este libro, pero puede entenderse como que la gráfica de la curva tiene área igual a cero. En la práctica, la mayor parte de las curvas y, desde luego, todas las que van a aparecer en este texto, cumplen dicha propiedad, por lo que no insistiremos más sobre esta cuestión.

La función de distribución anterior verifica las condiciones de la definición 5.10 y existe, por tanto, la correspondiente función de densidad que se obtiene derivando F respecto de x y de y (o respecto de y y de x). Así,

$$f(x,y) = \begin{cases} 1 & \text{si } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Como puede observarse, la segunda derivada parcial de la función de distribución, $\frac{\partial^2 F(x,y)}{\partial x \partial y}$, no existe para los puntos (x,y) que se encuentran en el borde del cuadrado unidad, esto es, en los lados del cuadrado. En dicha curva no existirá, por tanto, la función de densidad. ◁

Las funciones de densidad y de distribución de una variable aleatoria bidimensional continua están relacionadas; se puede obtener una de ellas a partir de la otra.

Proposición 5.7 *Sea (X,Y) una variable aleatoria bidimensional continua y sean F y f sus funciones de distribución y de densidad, respectivamente. Entonces,*

1. $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$ *para todo* $(x,y) \in \mathbb{R}^2$.
2. $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt ds$ *para todo* $(x,y) \in \mathbb{R}^2$.

Demostración. La primera de las dos expresiones es la definición de función de densidad conjunta.

Probemos el segundo resultado, integrando iteradamente la función de densidad conjunta. De esta forma,

$$\begin{aligned} \int_{s=-\infty}^{x} \int_{t=-\infty}^{y} f(s,t)\, dt\, ds &= \int_{s=-\infty}^{x} \int_{t=-\infty}^{y} \frac{\partial^2 F(s,t)}{\partial s \partial t}\, dt\, ds = \\ &= \int_{s=-\infty}^{x} \int_{t=-\infty}^{y} \frac{\partial}{\partial t} \left[\frac{\partial F(s,t)}{\partial s} \right] dt\, ds = \\ &= \int_{s=-\infty}^{x} \left[\frac{\partial F(s,t)}{\partial s} \right]_{t=-\infty}^{y} ds = \\ &= \int_{s=-\infty}^{x} \frac{\partial F(s,y)}{\partial s}\, ds = \\ &= F(s,y)|_{s=-\infty}^{x} = F(x,y)\,, \end{aligned}$$

resultado obtenido sin más que tener en cuenta que $\frac{\partial F(s,t)}{\partial s}$ es continua respecto de t (por ser derivable) y que $F(x,-\infty) = F(-\infty,y) = 0$. ♣

Observe el lector que resultados análogos relacionaban la función de densidad y de distribución de una variable aleatoria unidimensional continua (proposición 4.9) y las funciones de probabilidad y de distribución de una variable aleatoria bidimensional discreta (proposición 5.4).

Veamos otras propiedades de interés de las variables continuas bidimensionales.

Proposición 5.8 *Sea (X,Y) una variable aleatoria bidimensional continua y sea f su función de densidad. Entonces:*

1. *Para cualesquiera a,b,c y d, números reales tales que $a<b$ y $c<d$,*

$$p[a<X\leq b,\ c<Y\leq d]=\int_a^b\int_c^d f(x,y)\,dydx\ .$$

2. *Para cualquier subconjunto B de la σ-álgebra de Borel del plano $\mathbb{R}^2$,*

$$p[(X,Y)\in B]=\int\int_B f(x,y)\,dxdy\ .$$

3. *$f(x,y)\geq 0$ para todo $(x,y)\in\mathbb{R}^2$.*

Demostración. En primer lugar, probemos la propiedad *1*, sin más que tener en cuenta las propiedades de los límites de integración en las integrales dobles:

$$\begin{aligned}
p[a<X\leq b,c<Y\leq d] &= F(b,d)-F(b,c)-F(a,d)+F(a,c)=\\
&=\int_{-\infty}^b\int_{-\infty}^d f(x,y)dydx-\int_{-\infty}^b\int_{-\infty}^c f(x,y)dydx-\\
&\quad-\int_{-\infty}^a\int_{-\infty}^d f(x,y)dydx+\int_{-\infty}^a\int_{-\infty}^c f(x,y)dydx=\\
&=\int_{-\infty}^b\left[\int_{-\infty}^d f(x,y)dy-\int_{-\infty}^c f(x,y)dy\right]dx-\\
&\quad-\int_{-\infty}^a\left[\int_{-\infty}^d f(x,y)dy-\int_{-\infty}^c f(x,y)dy\right]dx=\\
&=\int_{-\infty}^b\left[\int_c^d f(x,y)dy\right]dx-\int_{-\infty}^a\left[\int_c^d f(x,y)dy\right]dx=\\
&=\int_a^b\int_c^d f(x,y)dydx\ .
\end{aligned}$$

Este resultado afirma que la probabilidad de un rectángulo del plano $\mathbb{R}^2$ se obtiene integrando en dicho rectángulo la función de densidad. La propiedad *2* (en este libro no desarrollaremos su demostración) es una generalización del mismo.

Finalmente, probemos la propiedad *3*. Si en un punto (x_0,y_0), donde la función de densidad es continua, ésta tomase un valor negativo, la continuidad de la función obligaría a que fuese negativa en un rectángulo $\mathcal{R}$ alrededor de (x_0,y_0), esto es, se cumpliría

$$f(x,y)<0\quad\text{para todo}\quad(x,y)\in\mathcal{R}\ .$$

Por tanto, la probabilidad del rectángulo $\mathcal{R}$, por la propiedad *1*, sería

$$p[(X,Y)\in\mathcal{R}]=\int\int_{\mathcal{R}} f(x,y)dxdy<0\ ,$$

ya que la integral de una función negativa es también negativa.

Puesto que es un absurdo que una probabilidad sea negativa, debemos concluir que la función de densidad conjunta es positiva en todos los puntos en que es continua.

En los puntos de discontinuidad de f podemos asignar a la función valores arbitrarios, como hemos comentado; por coherencia, se asignan valores no negativos. ♣

OBSERVACIONES

1) La tercera condición de la definición 5.10 garantiza que la probabilidad no se concentre en un subespacio unidimensional del plano $\mathbb{R}^2$. En el caso de las variables aleatorias unidimensionales, dicha condición estaba implícita en la definición por lo que no era necesario exigirla. En ese caso, la continuidad de la función de distribución permitía asegurar que la probabilidad no se concentrase en un punto. En el caso bidimensional (y, en general, en el caso n-dimensional), esa condición hay que establecerla explícitamente. Veamos el siguiente ejemplo.

Ejemplo 5.8. Se lanza una moneda equilibrada. Si sale cara, se eligen al azar y separadamente dos números, X e Y, entre 0 y 1; si sale cruz, se elige un solo número, X, y se toma $Y = X$. Hallemos la función de distribución conjunta de (X, Y).

Llamemos c al suceso *obtener cara* y z al suceso *obtener cruz*. Entonces,

$$\begin{aligned} F(x,y) &= p[X \le x, Y \le y] = \\ &= p[X \le x, Y \le y \mid c] \cdot p(c) + p[X \le x, Y \le y \mid z] \cdot p(z) = \\ &= p[X \le x, Y \le y \mid c] \cdot 0.5 + p[X \le x, Y \le y \mid z] \cdot 0.5 \, . \end{aligned}$$

Por tanto, si $0 < x < 1$, $0 < y < 1$,

$$p[X \le x, Y \le y \mid c] = x \cdot y \, ,$$

es decir, si sale cara, la probabilidad se reparte uniformemente en el cuadrado de lado unidad, $(0,1) \times (0,1)$, de la misma manera que en el ejemplo 5.7, y

$$p[X \le x, Y \le y \mid z] = \min(x, y) \, ,$$

esto es, si sale cruz, la probabilidad se reparte uniformemente en la diagonal del cuadrado unidad anterior. Es decir, si $0 < x < 1$, $0 < y < 1$,

$$F(x,y) = 0.5 \cdot x \cdot y + 0.5 \cdot \min(x, y) \, .$$

De igual forma, puede el lector hallar los valores de F para el resto de los casos, resultando:

$$F(x,y) = \begin{cases} 0 & x \le 0 \quad \text{o} \quad y \le 0 \\ 0.5 \cdot x \cdot y + 0.5 \cdot \min(x,y) & 0 < x < 1, \;\; 0 < y < 1 \\ x & 0 < x < 1, \;\; y \ge 1 \\ y & 0 < y < 1, \;\; x \ge 1 \\ 1 & x \ge 1, \;\; y \ge 1 \end{cases}$$

Derivando esta función respecto de x y respecto de y, obtenemos:

$$f(x,y) = \begin{cases} 0.5 & 0 < x < 1, \;\; 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

cuya integral en el cuadrado unidad, $(0,1) \times (0,1)$, resulta ser distinta de la unidad. Por tanto, (X, Y) no es una variable aleatoria bidimensional continua. El hecho es que la mitad de la probabilidad se reparte uniformemente en el cuadrado unidad y la otra mitad en la diagonal del mismo. ◁

2) De la propiedad *2* de la última proposición se deduce que la probabilidad de que una variable aleatoria bidimensional, (X,Y), esté en una curva del plano (de área cero, como comentamos anteriormente) es nula, ya que la integral de una función en una curva de ese tipo es cero. Por ejemplo, serán nulas las probabilidades $p[Y = X^2]$, $p[Y = X]$ o $p[X = 2]$, es decir, las probabilidades de que (X,Y) tome valores en la parábola del plano $y = x^2$, en la bisectriz del primer y tercer cuadrante, $y = x$, o en la recta vertical $x = 2$, respectivamente. Asimismo, las probabilidades de los conjuntos de Borel del plano serán iguales, tanto si incluyen a los *bordes*, como si no; en concreto, la exclusión de todos o alguno de los lados de un rectángulo no afectará a su probabilidad.

3) A diferencia de la función de probabilidad, la función de densidad conjunta no expresa la probabilidad de los puntos del plano $\mathbb{R}^2$ (de hecho, como acabamos de ver, estas probabilidades son siempre nulas). Como ocurría para las variables unidimensionales $f(x,y)$ describe la densidad de probabilidad en torno al punto (x,y). Siguiendo un razonamiento análogo al de la observación **3** del apartado 4.5, obtenemos el siguiente resultado:

$$\lim_{\substack{h\to 0\\ k\to 0}} \frac{p\left[x-\frac{h}{2}\le X\le x+\frac{h}{2}\ ,\ y-\frac{k}{2}\le Y\le y+\frac{k}{2}\right]}{h\cdot k} = f(x,y)\ ,$$

es decir, $f(x,y)$ es la probabilidad por unidad de área en un rectángulo *infinitesimal* alrededor del punto (x,y).

Así, valores grandes de la función de densidad indicarán zonas de puntos con mucha probabilidad (ello se corresponderá con crecimientos notables de la función de distribución, $F(x,y)$); en cambio, valores pequeños de $f(x,y)$ indicarán zonas de puntos con poca probabilidad (crecimientos suaves de $F(x,y)$).

La propiedad *3* de la proposición anterior, junto con el hecho de que la integral de la función de densidad conjunta en todo el plano es igual a la unidad, caracterizan una función de densidad. En concreto, puede probarse el resultado siguiente.

Proposición 5.9 *Sea g una función de $\mathbb{R}^2$ en $\mathbb{R}$ continua salvo, a lo sumo, en un número finito de curvas. Entonces, g es la función de densidad de una variable aleatoria bidimensional continua si y sólo si cumple:*

1. $g(x,y) \ge 0$ *para todo* $(x,y) \in \mathbb{R}^2$.
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dxdy = 1$.

La demostración de la condición necesaria, que no realizaremos, es similar a la de la proposición 4.11.

Como ocurre para las funciones de probabilidad marginal cuando la variable bidimensional es discreta, si (X,Y) es continua, se pueden obtener las funciones de densidad marginales a partir de la conjunta. El siguiente resultado indica cómo hacerlo.

Proposición 5.10 *Si (X,Y) es una variable aleatoria bidimensional continua, entonces X e Y son variables unidimensionales continuas y, además,*

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \,.$$

Demostración. En efecto,

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \lim_{y\to\infty} \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\, dtds \,.$$

En consecuencia,

$$F_X(x) = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(s,t)\, dt\right] ds \,.$$

Por otra parte,

$$F_X(x) = \int_{-\infty}^{x} f_X(s)ds \,.$$

Por tanto, $\int_{-\infty}^{\infty} f(s,t)dt$ es la función de densidad de X (el lector puede comprobar que dicha función verifica las condiciones que caracterizan a la función de densidad de una variable aleatoria continua). De forma análoga, demostraríamos el resultado para la variable Y. ♣

Por tanto, si (X,Y) es una variable aleatoria bidimensional continua, sus componentes, X e Y, son variables aleatorias unidimensionales continuas. El resultado recíproco no es cierto, esto es, si X e Y son variables continuas, no tiene por qué serlo la variable bidimensional (X,Y). El siguiente ejemplo, de desarrollo analítico algo complicado, ilustra este comentario. Al lector poco familiarizado con el cálculo en varias variables le costará seguir su desarrollo. Pero debajo de este ejemplo subyace una idea que tal vez sea fácil de entender. En él estudiamos la variable bidimensional (X,X^2) para una variable continua X. Como veremos más adelante, la función de densidad conjunta de ambas variables es distinta de cero sólo para los valores que toma la variable aleatoria bidimensional (X,X^2), esto es, para los puntos del plano que se encuentran en la curva $y = x^2$. Entonces, las probabilidades de cualquier suceso (que, como hemos visto, son las integrales de la función de densidad conjunta en éstos) serían todas iguales a cero.

Ello descarta que las probabilidades para la variable aleatoria (X,X^2) puedan obtenerse integrando una función de densidad conjunta y, en consecuencia, que (X,X^2) sea continua. En definitiva, las variables aleatorias continuas tienen que *llenar* subconjuntos del plano de área no nula y no encontrarse, con probabilidad distinta de cero, en curvas de área igual a cero.

Veamos el ejemplo.

Ejemplo 5.9. Sea X una variable aleatoria continua, con función de densidad

$$f_X(x) = \begin{cases} 1 & \text{si } 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

y con función de distribución

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Sea $Y = X^2$; entonces, Y es continua. Haciendo un cambio de variable, puede comprobarse que su función de densidad (que no utilizaremos en este ejemplo) vale

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{si } 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Obtengamos la función de distribución conjunta de X e Y. Como X e Y varían en el intervalo $(0,1)$, $F(x,y)$ valdrá cero si x o y son negativos, esto es, en los cuadrantes segundo, tercero y cuarto del plano. Por el mismo motivo, si x e y son mayores que la unidad, $F(x,y) = 1$.

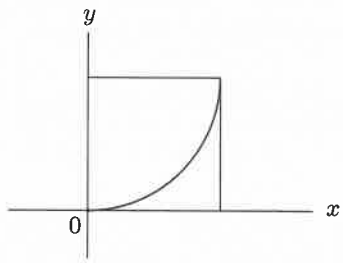


FIG. 5.12.

Sea (x,y) un punto con $0 \leq x \leq 1$ e $y \geq x^2$. Entonces, los sucesos $[X \leq x, X^2 \leq y]$ y $[X \leq x]$ son iguales, ya que, si $X \leq x$, entonces $X^2 \leq x^2 \leq y$. Por tanto,

$$\begin{aligned} F(x,y) &= p[X \leq x, Y \leq y] = p[X \leq x, X^2 \leq y] = \\ &= p[X \leq x] = F_X(x) = x \,. \end{aligned}$$

Análogamente, si (x,y) es un punto con $0 \leq y \leq 1$ e $y < x^2$, el suceso $[X \leq x, X^2 \leq y]$ es igual al suceso $[X^2 \leq y]$, ya que si $X^2 \leq y$, entonces $X^2 \leq y < x^2$. Por tanto,

$$\begin{aligned} F(x,y) &= p[X \leq x, Y \leq y] = p[X \leq x, X^2 \leq y] = p[X^2 \leq y] = \\ &= p[X \leq \sqrt{y}] = F_X(\sqrt{y}) = \sqrt{y} \,. \end{aligned}$$

En definitiva, la función de distribución conjunta de (X,Y) vale

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ o } y < 0 \\ x & \text{si } 0 \leq x \leq 1,\ y \geq x^2 \\ \sqrt{y} & \text{si } 0 \leq y \leq 1,\ y < x^2 \\ 1 & \text{si } x > 1,\ y > 1 \end{cases}$$

Puede comprobarse que para esta función de distribución no existe la derivada parcial con respecto a x en todos los puntos. Por ejemplo, si $y = 0.16$, la función $F(x, 0.16)$ vale

$$F(x, 0.16) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 0.4 \\ 0.4 & \text{si } x > 0.4 \end{cases}$$

que no es derivable con respecto a x en el punto $x = 0.4$.

En definitiva, la función de distribución de (X, Y) no cumple las condiciones exigidas en la definición de variable aleatoria continua y, por tanto, (X, Y) no es una variable aleatoria bidimensional continua. ◁

Teniendo en cuenta que la integral doble de una función positiva en una cierta región del plano es el volumen comprendido entre dicha región y la superficie que representa la función, podemos *visualizar* algunos de los resultados anteriores sobre un sistema tridimensional de ejes cartesianos. De este modo, la representación de la función de densidad conjunta será una superficie sobre el plano xy que deja debajo un volumen igual a la unidad, siendo el volumen que queda por *encima* el rectángulo $(a, b) \times (c, d)$, la probabilidad del suceso $[a < X < b, c < X < d]$.

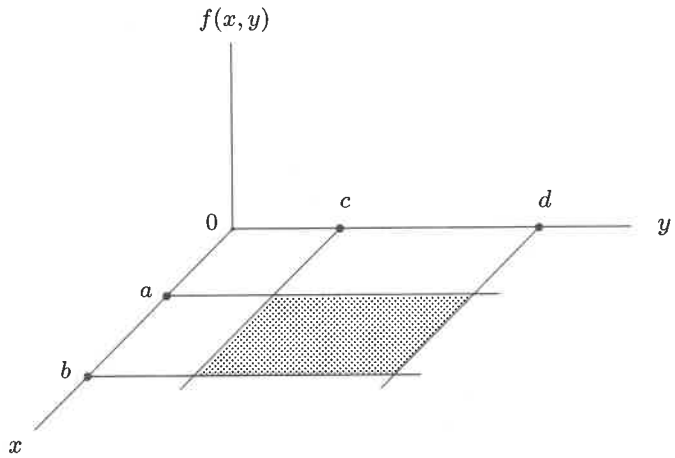


FIG. 5.13.

Ejemplo 5.10. Consideremos de nuevo la variable aleatoria propuesta en el ejemplo 5.7 cuya función de densidad es:

$$f(x, y) = \begin{cases} 1 & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

La representación gráfica de dicha función se presenta en la figura 5.14. Por tanto, se trata de una superficie plana elevada sobre el cuadrado unidad del plano xy (evidentemente, fuera de ese cuadrado $f(x, y)$ vale cero). Advierta el lector que la sencillez de

la función deja ver con claridad que *debajo* de esta superficie hay un volumen igual a la unidad, en concreto, un cubo de lado uno.

A partir de esta función de densidad es posible calcular probabilidades referidas a la variable (X, Y), sin más que integrar dicha función en el recinto correspondiente. Por ejemplo,

$$p\left[\frac{1}{2} < X < 1,\ \frac{1}{4} < Y < \frac{1}{2}\right] = \int_{x=\frac{1}{2}}^{1} \int_{y=\frac{1}{4}}^{\frac{1}{2}} 1\, dydx = \frac{1}{8}\,.$$

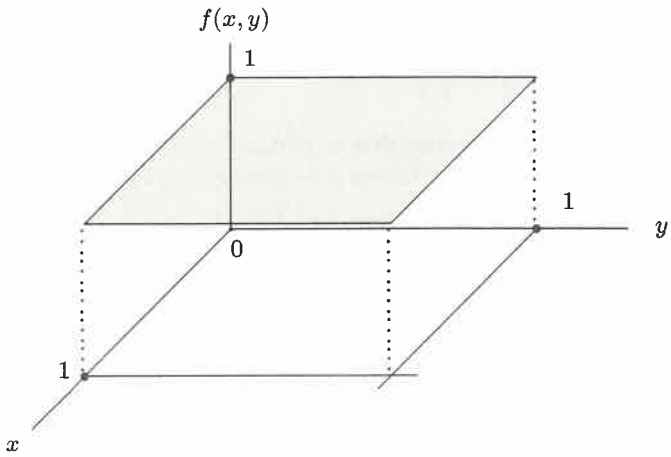


FIG. 5.14.

Por otro lado, la probabilidad de que la variable X tome valores mayores que la variable Y, cuyo recinto aparece en la figura 5.15 es

$$p[X > Y] = \int_{x=0}^{1} \int_{y=0}^{x} 1\, dydx = \int_{x=0}^{1} xdx = \frac{1}{2}\,.$$

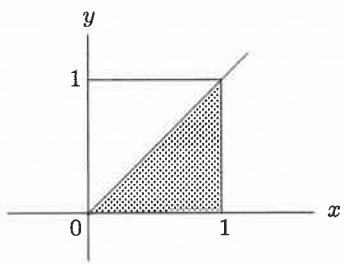


FIG. 5.15.

Estas probabilidades pueden ser calculadas con facilidad mediante procedimientos geométricos.

Puede el lector, de igual forma, hallar la función de distribución de (X, Y) mediante la función de densidad conjunta, integrando en el rectángulo $(-\infty, x] \times (-\infty, y]$ o calculando el volumen que queda encima del mismo.

Por otra parte, la proposición 5.10 nos permite obtener las funciones de densidad marginales:

$$f_X(x) = \begin{cases} \int_{x=0}^{1} 1\, dy = 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

y, análogamente,

$$f_Y(y) = \begin{cases} \int_{y=0}^{1} 1\, dx = 1 & 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Compruebe el lector que, tomando límites adecuadamente en $F(x, y)$, hallaría $F_X(x)$ y $F_Y(y)$, y derivando, obtendría las mismas funciones de densidad marginales. ◁

Todo lo anterior se puede generalizar para variables n-dimensionales.

Definición 5.12 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional y F su función de distribución. Decimos que $(X_1, \ldots, X_n)$ es una variable aleatoria n-dimensional continua si:*

1. *F es una función continua en cada variable.*
2. *Existen sus primeras derivadas parciales, $\frac{\partial}{\partial x_i} F(x_1, \ldots, x_i, \ldots, x_n)$ para $i = 1, \ldots, n$. Asimismo, su n-ésima derivada respecto de todas las variables, $\frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(x_1, \ldots, x_n)$, existe y es continua salvo, tal vez, a lo largo de un número finito de curvas en $\mathbb{R}^n$.*
3. $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1 .$$

Para facilitar el cálculo de probabilidades, se define *la función de densidad conjunta como una función,*

$$f : \mathbb{R}^n \to \mathbb{R} ,$$

tal que

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(x_1, \ldots, x_n)$$

para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Si fuera necesario para evitar confusiones, incluiríamos subíndices en las notaciones de la función de distribución y de densidad.

La función de densidad conjunta cumple las siguientes propiedades:

1. $F(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f(s_1,\ldots,s_n)ds_n\ldots ds_1$.
2. Si B es un subconjunto de la σ-álgebra de Borel en $\mathbb{R}^n$

$$p[(X_1,\ldots,X_n)\in B] = \int\cdots\int_B f(x_1,\ldots,x_n)dx_1\ldots dx_n\ .$$

3. $f(x_1,\ldots,x_n)\geq 0$ para todo $(x_1,\ldots,x_n)\in\mathbb{R}^n$.
4. La marginal de la i-ésima variable, $f_{X_i}(x_i)$, se calcula integrando

$$\int_{x_1}\cdots\int_{x_{i-1}}\int_{x_{i+1}}\cdots\int_{x_n} f(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n$$

para todo $i=1,\ldots,n$.

La tercera propiedad y el hecho de que la integral en todo el espacio sea la unidad, caracterizan a la función de densidad conjunta.

5.5 Distribuciones condicionadas

Sea (X,Y) una variable aleatoria bidimensional. Como ya sabemos, el estudio de su comportamiento probabilístico individual se hace a partir de sus distribuciones marginales (función de distribución marginal y función de densidad o de probabilidad marginal). No obstante, y como ya hemos visto, el estudio de probabilidades de sucesos referentes simultáneamente a ambas variables exige el conocimiento de su distribución conjunta. Por ejemplo, si queremos calcular la probabilidad del suceso

$$[X\leq x, Y\leq y] = \{\omega\in\Omega \mid X(\omega)\leq x \text{ e } Y(\omega)\leq y\},$$

podemos hacerlo si conocemos la función de distribución conjunta F (de hecho, vale $F(x,y)$). Aunque el suceso puede descomponerse en función de sucesos referentes a cada una de las dos variables aleatorias en la forma

$$[X\leq x, Y\leq y] = [X\leq x]\cap[Y\leq y]\ ,$$

su probabilidad no puede, en general, obtenerse a partir de la de ambos sucesos porque, como hemos visto, el cálculo de probabilidades de la intersección de dos sucesos exige más información que la que suministran las probabilidades de los sucesos que se interseccionan. La información adicional necesaria se refiere a la relación entre ambas variables y puede suministrarse de varias maneras: bien a través de su distribución conjunta (como apuntamos anteriormente) o bien, según veremos en esta sección, a través de las llamadas distribuciones condicionadas.

Otro enfoque de este mismo problema es el siguiente: en ausencia de información sobre una de las variables aleatorias, por ejemplo Y, el comportamiento probabilístico de la otra, X, viene dado por su distribución marginal. Pero, a veces, se dispone de información sobre el valor que ha tomado la variable Y, información completa (esto es, se sabe que ha ocurrido el suceso $[Y=y]$), o únicamente

parcial (se conoce que Y ha tomado un valor perteneciente a un cierto subconjunto, B, de la recta o, dicho de otra forma, se sabe que ha ocurrido el suceso $[Y \in B]$). En cualquier caso, esta información favorece a unos puntos de Ω sobre otros y, en consecuencia, modifica las probabilidades relativas a la variable X. La distribución de probabilidades de X resultante recibirá el nombre de **distribución condicionada** por la información disponible, esto es, por el suceso referente a Y que sabemos que ha ocurrido. El siguiente ejemplo ayuda a ilustrar estas ideas.

Ejemplo 5.11. Supongamos que se lanzan dos dados perfectos de forma independiente. Sea X la variable aleatoria que recoge *el resultado del primer dado*, e Y el *menor resultado obtenido entre los dos dados*. (X, Y) es una variable aleatoria bidimensional discreta, cuya función de probabilidad conjunta, presentada en la siguiente tabla de doble entrada, fue obtenida en el ejemplo 5.5.

$X \setminus Y$	1	2	3	4	5	6	
1	6/36	0	0	0	0	0	1/6
2	1/36	5/36	0	0	0	0	1/6
3	1/36	1/36	4/36	0	0	0	1/6
4	1/36	1/36	1/36	3/36	0	0	1/6
5	1/36	1/36	1/36	1/36	2/36	0	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	11/36	9/36	7/36	5/36	3/36	1/36	1

La función de probabilidad marginal de Y, que aparece en la última fila, indica las probabilidades de cada uno de los valores posibles de Y en ausencia de información adicional sobre los resultados del experimento.

Supongamos que una persona se encuentra presente en los lanzamientos y observa que el resultado del primero de ellos es un tres, esto es, que ha ocurrido el suceso $[X = 3]$. Los posibles resultados del experimento ya no serán los 36 pares que forman el espacio muestral,

$$\Omega = \{(i, j); i = 1, \ldots, 6,\ j = 1, \ldots, 6\}\ ,$$

sino que quedarán reducidos a los seis resultados del suceso

$$[X = 3] = \{(3, j); j = 1, \ldots, 6\}.$$

Como la variable aleatoria Y era el menor resultado en ambos lanzamientos, sus probabilidades cambian notablemente. Obsérvese que Y sólo toma ahora los valores 1, 2 y 3 (o si el lector prefiere no cambiar el conjunto de valores de Y, puede decir que toma los valores 1, 2, 3, 4, 5 y 6, pero que ahora las probabilidades de los valores 4, 5 y 6 son nulas).

Además, resulta razonable modificar las probabilidades de los sucesos $[Y = 1]$, $[Y = 2]$ e $[Y = 3]$. En ausencia de información, las probabilidades de esos sucesos eran, respectivamente, $\frac{11}{36}$, $\frac{9}{36}$ y $\frac{7}{36}$. Pero, de los seis pares que ahora son posibles, para cuatro de ellos (los pares $(3,3)$, $(3,4)$, $(3,5)$ y $(3,6)$) el mínimo es igual a tres, para uno de ellos (el par $(3,2)$) el mínimo es igual a dos, y para el par $(3,1)$ el mínimo es igual a uno. Como los resultados eran equiprobables, las probabilidades de los sucesos $[Y = 1]$, $[Y = 2]$ e $[Y = 3]$ estarán en las proporciones 1, 1 y 4 y, por tanto, valdrán, respectivamente, $\frac{1}{6}$, $\frac{1}{6}$ y $\frac{4}{6}$.

Este resultado hubiera podido obtenerse con ayuda de las probabilidades condicionadas que estudiamos en el capítulo 3, ya que las funciones de probabilidad no son sino probabilidades de sucesos. Las probabilidades que estamos buscando son los valores

$$p[Y = y \mid X = 3] \quad \text{para todo} \quad y = 1, \ldots, 6 \,,$$

y, por tanto,

$$p[Y = y \mid X = 3] = \frac{p[Y = y, X = 3]}{p[X = 3]} \,.$$

En concreto,

$$p[Y = 1 \mid X = 3] = \frac{p[Y = 1, X = 3]}{p[X = 3]} = \frac{1/36}{1/6} = \frac{1}{6} \,,$$

y el resto de los valores resultan ser:

$$\begin{aligned} p[Y = 2 \mid X = 3] &= \frac{1}{6} \\ p[Y = 3 \mid X = 3] &= \frac{4}{6} \qquad \text{y} \\ p[Y = 4 \mid X = 3] &= p[Y = 5 \mid X = 3] = p[Y = 6 \mid X = 3] = 0 \,, \end{aligned}$$

ya que el numerador de la última expresión es igual a cero para $y = 4, 5, 6$. ◁

Como muestra el ejemplo anterior, las distribuciones condicionadas resultan sencillas de hallar cuando la variable aleatoria bidimensional es discreta, ya que su obtención es un problema de cálculo de probabilidades condicionadas que ya estudiamos en el capítulo 3. Sin embargo, cuando la variable bidimensional es continua y se dispone de información completa sobre una de las variables (esto es, se conoce qué valor ha tomado), el estudio de las distribuciones condicionadas cambia de naturaleza según veremos más adelante.

5.5.1 Distribuciones discretas condicionadas

Supongamos que se dispone de una variable aleatoria bidimensional discreta (X, Y). En esta sección queremos estudiar la distribución de una de las variables cuando se sabe que ha ocurrido un suceso.

En la mayor parte de las aplicaciones prácticas, el suceso conocido es el valor que toma la otra variable; por ello, nos dedicaremos fundamentalmente a dicho problema. No obstante, como comentaremos al final de esta sección, puede condicionarse una distribución por sucesos diversos, dependiendo de la información disponible.

Los desarrollos posteriores se realizarán, indistintamente, para variables que toman un número finito, o infinito numerable, de valores. Comencemos definiendo lo que se entiende por una distribución condicionada.

Definición 5.13 *Sea (X, Y) una variable aleatoria bidimensional discreta y sea $f(x, y)$ su función de probabilidad.*

Si y_j es un valor tal que $p[Y = y_j] \neq 0$, definimos la función de probabilidad de X condicionada por $Y = y_j$ como la aplicación,

$$f_{(X \mid Y = y_j)} : I\!R \to I\!R \,,$$

tal que

$$f_{(X \mid Y=y_j)}(x) = p[X = x \mid Y = y_j] \quad \textit{para todo} \quad x \in \mathbb{R} .$$

Análogamente, si $p[X = x_i] \neq 0$, *definimos la función de probabilidad de* Y *condicionada por* $X = x_i$ *como la aplicación,*

$$f_{(Y \mid X=x_i)} : \mathbb{R} \to \mathbb{R} ,$$

tal que,

$$f_{(Y \mid X=x_i)}(y) = p[Y = y \mid X = x_i] \quad \textit{para todo} \quad y \in \mathbb{R} .$$

La función de probabilidad condicionada es, realmente, una función de probabilidad. En primer lugar, $p[X = x_i \mid Y = y_j]$ es no negativa, como cociente de probabilidades, que son cantidades no negativas. Por otro lado,

$$\sum_{i=1}^{\infty} f_{(X \mid Y=y_j)}(x_i) = \sum_{i=1}^{\infty} p[X = x_i \mid Y = y_j] = \sum_{i=1}^{\infty} \frac{p[X = x_i, Y = y_j]}{p[Y = y_j]} = \frac{\sum_{i=1}^{\infty} p[X = x_i, Y = y_j]}{p[Y = y_j]} ,$$

que resulta ser la unidad, ya que numerador y denominador coinciden.

OBSERVACIONES

1) Según muestra la definición, el cálculo de la función de probabilidad condicionada exige únicamente el conocimiento de la conjunta. En efecto, si x_i es un valor de la variable X,

$$f_{(X \mid Y=y_j)}(x_i) = p[X = x_i \mid Y = y_j] = \frac{p[X = x_i, Y = y_j]}{p[Y = y_j]} = \frac{f(x_i, y_j)}{f_Y(y_j)}$$

donde, como sabemos, el denominador se calcula:

$$f_Y(y_j) = \sum_{i=1}^{\infty} p[X = x_i, Y = y_j] = \sum_{i=1}^{\infty} f(x_i, y_j) .$$

2) La exigencia de que la probabilidad de $[Y = y_j]$ sea no nula no es gratuita. Si fuese cero, la expresión de $f_{(X \mid Y=y_j)}(x)$ no tendría sentido (se trataría de una división por cero) y su concepto carecería de lógica (se condicionaría por un suceso que no puede ocurrir).

3) La definición de función de probabilidad condicionada para variables discretas, recuerda las ideas que se planteaban en el estudio de las probabilidades condicionadas. La función de probabilidad de X condicionada por $[Y = y_j]$ en un punto x_i es la probabilidad del suceso $[X = x_i]$ condicionada por $[Y = y_j]$; y dicha probabilidad tiene que ver con la ocurrencia del primero de los sucesos dentro del segundo y, por tanto, con la intersección de ambos. La división por $p[Y = y_j]$ sólo tiene por objeto normalizar la distribución condicionada, esto es, que la suma de los valores no nulos de la función de probabilidad sea igual a la unidad.

4) La definición anterior puede escribirse de otra forma; a saber, *si* y_j *es un valor de* Y *tal que* $p[Y = y_j] > 0$, *entonces*

$$f(x_i, y_j) = f_Y(y_j) \cdot f_{(X \mid Y=y_j)}(x_i) \, .$$

Esta expresión resulta útil cuando las variables X e Y se generan en fases sucesivas del experimento, de forma que el resultado de Y determina cuál es el contexto en el que continúa el experimento hasta la obtención de un valor de la variable X. En estas situaciones suele conocerse la distribución de probabilidades marginal de Y y las distribuciones de probabilidad de la variable X en los distintos contextos en los que pueda obtenerse su valor. Veamos un ejemplo.

Ejemplo 5.12. Supongamos que un auditor selecciona un cierto número, X, de facturas de un archivador; escoge ese número al azar entre 5 y 8. Habiendo elegido un número x de facturas, el tiempo en minutos, Y, que tarda en revisarlas se ajusta a la función de probabilidad

$$p_x[Y = y] = \frac{10 - x}{x} \cdot \left(\frac{x}{10}\right)^y \quad \text{para } y = 1, 2, \ldots$$

En este caso, conocemos la función de probabilidad marginal de X que, puesto que el número se escoge al azar entre 5 y 8, podemos describir como

$$f_X(x) = \begin{cases} \frac{1}{4} & \text{si } x = 5, 6, 7, 8 \\ 0 & \text{en el resto} \end{cases}$$

Conocemos, también, la función de probabilidad de Y condicionada por $[X = x]$ para $x = 5, 6, 7, 8$, función de probabilidad que valdrá

$$f_{(Y \mid X=x)}(y) = p_x[Y = y]$$

donde p_x es la probabilidad definida al comienzo del ejemplo.

La función de probabilidad conjunta puede calcularse como

$$f(x, y) = f_X(x) \cdot f_{(Y \mid X=x)}(y) = \frac{1}{4} \cdot \frac{10 - x}{x} \cdot \left(\frac{x}{10}\right)^y$$

para $y = 1, 2, \ldots$, y $x = 5, 6, 7, 8$.

A partir de ella, puede, por ejemplo, obtenerse la función de probabilidad marginal de Y, esto es, la distribución de probabilidad del número de minutos necesario para revisar las facturas, sin consideración del número de facturas elegidas. Asimismo, se puede hallar la función de probabilidad de X condicionada a que $Y = y$, es decir, la distribución del número de facturas seleccionadas sabiendo que se ha tardado y minutos en revisarlas. ◁

Para las variables bidimensionales discretas pueden construirse funciones de probabilidad condicionadas por sucesos distintos de los hasta ahora considerados, esto es, sucesos de la forma $[Y \leq y]$ o de la forma $[X + Y \geq z]$, por ejemplo. Las definiciones no varían en lo esencial, aunque sí en su aspecto, de la que hemos dado cuando el suceso que condiciona es de la forma $[X = x_i]$ o $[Y = y_j]$. Demos la siguiente definición general.

Definición 5.14 *Sea (X,Y) una variable bidimensional discreta y sea B un subconjunto del plano, tal que $p[(X,Y) \in B] \neq 0$. Definimos la función de probabilidad de X condicionada por el suceso $[(X,Y) \in B]$ como*

$$f_{(X \mid (X,Y)\in B)}(x) = p[X = x \mid (X,Y) \in B] .$$

Como puede observarse, esta definición no incluye ninguna idea que sea esencialmente distinta de la sugerida en la definición 5.13, esto es, las funciones de probabilidad condicionadas se definen y calculan utilizando las técnicas referentes a las probabilidades condicionadas de sucesos. Ilustraremos esta observación con un ejemplo.

Ejemplo 5.13. Sea (X,Y) una variable bidimensional cuya función de probabilidad viene dada por la siguiente tabla de doble entrada.

$X\backslash Y$	1	2	3	
1	0.1	0.2	0.1	0.4
2	0.2	0.1	0	0.3
3	0.1	0.2	0	0.3
	0.4	0.5	0.1	1

Las distribuciones marginales aparecen en los bordes derecho e inferior de la tabla. Por ejemplo, la función de probabilidad marginal de la variable X tiene como valores no nulos $f_X(1) = 0.4$, $f_X(2) = 0.3$ y $f_X(3) = 0.3$.

Las funciones de probabilidad condicionadas por valores puntuales de una de las variables se calculan sin dificultad. Por ejemplo, la función de probabilidad de X condicionada por el suceso $[Y = 2]$ será

$$f_{(X \mid Y=2)}(x) = p[X = x \mid Y = 2] = \begin{cases} 0.4 & x = 1 \\ 0.2 & x = 2 \\ 0.4 & x = 3 \\ 0 & \text{en el resto} \end{cases}$$

Veamos probabilidades condicionadas por otro tipo de sucesos. Consideremos, por ejemplo, el suceso $[X \geq Y]$. Su probabilidad vale 0.7, como puede comprobar el lector fácilmente. La función de probabilidad de X, condicionada por el suceso $[X \geq Y]$, será, obviamente, nula salvo para los valores $x = 1, 2, 3$. Para ellos,

$$f_{(X \mid X\geq Y)}(1) = p[X = 1 \mid X \geq y] = \frac{p[X = 1, X \geq Y]}{p[X \geq Y]} = \frac{p[X = 1, Y = 1]}{p[X \geq Y]} = \frac{1}{7} .$$

Análogamente,

$$f_{(X \mid X\geq Y)}(2) = p[X = 2 \mid X \geq y] = \frac{p[X = 2, X \geq Y]}{p[X \geq Y]} = \frac{3}{7} ,$$

ya que el suceso $[X = 2, X \geq Y]$ se puede escribir como

$$[X = 2, X \geq Y] = [X = 2, Y = 1] \cup [X = 2, Y = 2]$$

y, por tanto, su probabilidad es

$$p[X = 2, X \geq Y] = p[X = 2, Y = 1] + p[X = 2, Y = 2] = 0.2 + 0.1 = \frac{3}{7} .$$

De la misma forma se obtiene

$$f_{(X \mid X \geq Y)}(3) = 0.3 \, .$$

Por tanto,

$$f_{(X \mid X \geq Y)} = \begin{cases} 1/7 & x = 1 \\ 3/7 & x = 2 \\ 3/7 & x = 3 \\ 0 & \text{en el resto} \end{cases}$$

función que, como se puede comprobar, es realmente de probabilidad.

Igualmente, calcularíamos la distribución de Y condicionada al suceso $[X + Y = 3]$. La probabilidad de dicho suceso es:

$$p[X + Y = 3] = p[X = 1, Y = 2] + p[X = 2, Y = 1] = 0.2 + 0.2 = 0.4 \, .$$

Entonces,

$$\begin{aligned} f_{(Y \mid X+Y=3)}(1) &= p[Y = 1 \mid X + Y = 3] = \\ &= \frac{p[Y = 1, X + Y = 3]}{p[X + Y = 3]} = \frac{p[X = 2, Y = 1]}{p[X + Y = 3]} = 0.5 \, . \end{aligned}$$

Análogamente hallaríamos $f_{(Y \mid X+Y=3)}(2)$, obteniéndose el mismo valor, 0.5. ◁

Advertirá el lector que, en todos los casos, podríamos definir la función de distribución condicionada,

$$F_{(X \mid (X,Y) \in B)}(x) = p[X \leq x \mid (X, Y) \in B]$$

y calcularla, bien directamente, bien acumulando la función de probabilidad. La escasa relevancia práctica de esta función hace que no insistamos en ella.

5.5.2 Distribuciones continuas condicionadas

Sea (X, Y) una variable aleatoria bidimensional continua. Supongamos, por analogía con el problema estudiado en la sección anterior, que ha ocurrido el suceso $[Y = y]$. Aunque, si dicho suceso ha ocurrido, es porque y es un valor posible de la variable Y, sabemos que su probabilidad es nula. Por tanto, no parece posible estudiar las funciones de densidad condicionadas como hicimos con las funciones de probabilidad condicionadas en variables discretas.

Pero, además, las funciones de densidad no son probabilidades, sino densidades de probabilidad, lo que constituye un motivo adicional para no poder aplicar a las funciones de densidad las técnicas de probabilidades condicionadas de sucesos que estudiamos en el capítulo 3.

Este segundo problema es resoluble. En efecto, la función de distribución sí es una probabilidad y, por tanto, podríamos plantearnos definir la función de distribución de X condicionada por el suceso $[Y = y]$ como

$$F_{(X \mid Y=y)}(x) = p[X \leq x \mid Y = y] \, .$$

Pero, como puede observarse, el problema esencial, que es el que hemos planteado en primer lugar, se mantiene; la probabilidad condicionada que aparece en la última expresión no existe y, por tanto, dicha expresión carece de sentido.

La solución a este problema consiste en definir de otra forma la función de distribución condicionada. La nueva definición no es arbitraria y, según veremos, conduce a soluciones razonables desde el punto de vista del cálculo de las distribuciones condicionadas.

Definición 5.15 *Sea (X,Y) una variable aleatoria bidimensional continua y sea y un valor posible de Y, es decir, tal que $f_Y(y) \neq 0$. Definimos la función de distribución de X condicionada por $Y = y$ como la aplicación,*

$$F_{(X \mid Y=y)} : \mathbb{R} \to \mathbb{R} ,$$

tal que

$$F_{(X \mid Y=y)}(x) = \lim_{h \to 0} p[X \leq x \mid y - h < Y \leq y] .$$

Obsérvese que, por su aspecto, la nueva definición recuerda mucho a la expresión que anteriormente hemos desechado. De hecho, estamos condicionando por el suceso $[y-h < Y \leq y]$ que en el límite cuando h tiende hacia cero es, precisamente, $[Y = y]$. Como

$$p[X \leq x \mid y - h < Y \leq y] = \frac{p[X \leq x, y - h < Y \leq y]}{p[y - h < Y \leq y]} ,$$

entonces

$$F_{(X \mid Y=y)}(x) = \frac{\lim_{h \to 0} p[X \leq x, y - h < Y \leq y]}{\lim_{h \to 0} p[y - h < Y \leq y]} .$$

Ahora bien, por la *continuidad* de las probabilidades a que aludíamos en el apartado 4.3,

$$\lim_{h \to 0} p[X \leq x, y - h < Y \leq y] = p[X \leq x, Y = y]$$

y

$$\lim_{h \to 0} p[y - h < Y \leq y] = p[Y = y] .$$

Así, la definición 5.15 conduce a un cociente cuyo numerador y denominador son iguales a cero. Sin embargo, el límite del cociente existe. En efecto,

$$\begin{aligned} F_{(X \mid Y=y)}(x) &= \lim_{h \to 0} \frac{p[X \leq x, y - h < Y \leq y]}{p[y - h < Y \leq y]} = \\ &= \lim_{h \to 0} \frac{F(x,y) - F(x, y-h)}{F_Y(y) - F_Y(y-h)} = \\ &= \lim_{h \to 0} \frac{(F(x,y) - F(x,y-h))/h}{(F_Y(y) - F_Y(y-h))/h} . \end{aligned}$$

Pero,

$$\lim_{h \to 0} \frac{F(x,y) - F(x,y-h)}{h} = \frac{\partial F(x,y)}{\partial y}$$

y

$$\lim_{h\to 0}\frac{F_Y(y)-F_Y(y-h)}{h}=\frac{dF_Y(y)}{dy}=f_Y(y)\,.$$

En definitiva, *la función de distribución de X condicionada por $Y=y$ es*

$$F_{(X\mid Y=y)}(x)=\frac{\frac{\partial F(x,y)}{\partial y}}{f_Y(y)}\,.$$

Comprobemos que la función obtenida es, realmente, una función de distribución. En primer lugar, la función es monótona creciente, dado que su derivada

$$\frac{dF_{(X\mid Y=y)}(x)}{dx}=\frac{\frac{\partial^2 F(x,y)}{\partial x\partial y}}{f_Y(y)}=\frac{f(x,y)}{f_Y(y)}\,,$$

es positiva por ser cociente de dos funciones que lo son.

Por otro lado,

$$F_{(X\mid Y=y)}(+\infty)=\frac{\frac{\partial F(+\infty,y)}{\partial y}}{f_Y(y)}=\frac{\frac{dF_Y(y)}{dy}}{f_Y(y)}=1$$

y

$$F_{(X\mid Y=y)}(-\infty)=\frac{\frac{\partial F(-\infty,y)}{\partial y}}{f_Y(y)}=\frac{0}{f_Y(y)}=0\,,$$

dado que $\frac{\partial F(x,y)}{\partial y}$ es continua respecto de x, al ser derivable.

Finalmente, $F_{(X\mid Y=y)}$ es continua respecto de x dado que es derivable.

Además, esta función de distribución de X condicionada a $Y=y$ lo es de una distribución continua, dado que es continua y su derivada, salvo un número finito de puntos, también lo es.

Se puede, por tanto, obtener *la función de densidad condicionada, $f_{(X\mid Y=y)}$, que será la derivada con respecto a x de la función de distribución condicionada. De esta forma,*

$$f_{(X\mid Y=y)}(x)=\frac{dF_{(X\mid Y=y)}(x)}{dx}=\frac{\frac{\partial^2 F(x,y)}{\partial x\partial y}}{f_Y(y)}=\frac{f(x,y)}{f_Y(y)}\,.$$

OBSERVACIONES

1) Nótese que la técnica de cálculo de las funciones de densidad condicionadas es análoga a la empleada para obtener las funciones de probabilidad condicionadas, sin más que sustituir en el numerador y en el denominador las funciones de probabilidad por funciones de densidad.

2) Obsérvese que, así como para variables discretas el suceso que condicionaba debía ser de probabilidad no nula, de forma que existieran las probabilidades condicionadas correspondientes, cuando la variable bidimensional es continua, la condición que se exige al suceso $[Y = y]$ es que $f_Y(y)$ sea distinto de cero. Ello permite que la función de densidad condicionada

$$f_{(X \mid Y=y)}(x) = \frac{f(x,y)}{f_Y(y)}$$

esté definida. Por tanto, podemos definir, en principio, tantas distribuciones de X condicionadas, como valores posibles de Y.

En cualquier caso, cuando se condiciona por sucesos de la forma $[Y = y]$, la condición que se exige, tanto para variables discretas como continuas, es que y sea un valor posible de la variable Y.

3) Al igual que en el caso discreto, podemos descomponer la función de densidad conjunta en producto de dos funciones de densidad: la marginal de Y y la de X condicionada por $[Y = y]$. Es decir, *si y es un valor de Y tal que $f_Y(y) > 0$, entonces*

$$f(x,y) = f_Y(y) \cdot f_{(X \mid Y=y)}(x) \, .$$

4) Cuando el suceso que condiciona la distribución tiene probabilidad no nula (lo que ocurrirá habitualmente si la información es incompleta), la función de distribución condicionada se obtiene con las técnicas de la probabilidad condicionada de sucesos. Así, *si B es un subconjunto de Borel del plano $\mathbb{R}^2$, tal que $p[(X,Y) \in B] \neq 0$, definimos la función de distribución de X condicionada por el suceso $[(X,Y) \in B]$ como*

$$F_{(X \mid (X,Y) \in B)}(x) = p[X \leq x \mid (X,Y) \in B] \, .$$

A partir de esta función, derivando, puede calcularse la función de densidad.

Es obvio que todas las formalizaciones que se han hecho para la distribución de X condicionada a valores de Y se pueden realizar al revés, es decir, para la distribución de Y condicionada por X.

Ejemplo 5.14. La variable X mide los ingresos anuales, en millones, de las familias de un estrato de la población, y la variable Y los gastos anuales en alimentos de las mismas (asimismo en millones). Asignamos a la variable aleatoria bidimensional (X,Y) la siguiente función de densidad:

$$f(x,y) = \begin{cases} \frac{1}{24}(x+y) & \text{si } 0 < x < 4, \ 0 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

Las funciones de densidad marginales se obtienen integrando esta función, y valen

$$f_X(x) = \begin{cases} \frac{x+1}{12} & \text{si } 0 < x < 4 \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \begin{cases} \frac{2+y}{6} & \text{si } 0 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

Obtengamos la distribución de probabilidades de los gastos anuales en alimentos, Y, de las familias con ingresos anuales iguales a 3 millones. Lo que se busca es, por tanto, la distribución de Y condicionada por el suceso $[X = 3]$. Vamos a calcular, en primer lugar, la función de densidad condicionada.

Para $0 < y < 2$, $f(3, y) = \frac{1}{24}(3 + y)$ y, por tanto,

$$f_{(Y \mid X=3)}(y) = \frac{f(3,y)}{f_X(3)} = \frac{\frac{1}{24}(3+y)}{\frac{3+1}{12}} = \frac{3+y}{8} .$$

Si y no está entre cero y dos, $f(3, y) = 0$ y, consecuentemente, $f_{(Y \mid X=3)}(y) = 0$. En definitiva,

$$f_{(Y \mid X=3)}(y) = \begin{cases} \frac{3+y}{8} & \text{si } 0 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

Esta función de densidad permite calcular probabilidades de diferentes intervalos de gasto en alimentos para las familias con tres millones de ingresos anuales. Así, la probabilidad de que una de estas familias tenga un gasto en alimentación entre uno y dos millones será:

$$p[1 \leq Y \leq 2 \text{ siendo } X = 3] = \int_1^2 f_{(Y \mid X=3)}(y)\, dy = \frac{9}{16} .$$

Obsérvese que una interpretación *literal* de la probabilidad pedida nos llevaría a plantear el cálculo de

$$p[1 \leq Y \leq 2 \mid X = 3] ,$$

probabilidad condicionada que no está definida, ya que al ser X continua, $p[X = 3]$ es cero.

Hallemos, ahora, la función de distribución condicionada. Podría calcularse directamente integrando la función de densidad condicionada que acabamos de obtener. No obstante, aunque, como verá el lector, los cálculos son bastante más complicados, aplicaremos la siguiente expresión, obtenida de la definición de función de distribución condicionada:

$$F_{(Y \mid X=3)}(y) = \frac{\frac{\partial F(x,y)}{\partial x}\Big|_{x=3}}{f_X(3)} .$$

Para ello, calculemos, en primer lugar, la función de distribución conjunta, integrando la función de densidad.

$$F(x,y) = \begin{cases} 0 & x \leq 0 \quad \text{o} \quad y \leq 0 \\ \frac{1}{48}(x^2y + xy^2) & 0 < x < 4, 0 < y < 2 \\ \frac{1}{12}(4y + y^2) & x \geq 4, 0 < y < 2 \\ \frac{1}{24}(x^2 + 2x) & 0 < x < 4, y \geq 2 \\ 1 & x \geq 4, y \geq 2 \end{cases}$$

Ahora ya estamos en condiciones de obtener la función de distribución condicionada, calculando las parciales correspondientes de la función de distribución conjunta. Entonces,

$$F_{(Y \mid X=3)}(y) = \begin{cases} 0 & \text{si } y < 0 \\ \frac{6y+y^2}{16} & \text{si } 0 \leq y \leq 2 \\ 1 & \text{si } y > 2 \end{cases}$$

Obviamente, derivando esta función resulta la función de densidad condicionada que obtuvimos anteriormente por procedimientos más sencillos.

Obsérvese que la distribución de Y condicionada por $[X = 3]$ es distinta a la distribución marginal de Y, es decir, el conocimiento de que X ha tomado el valor 3, cambia la distribución de probabilidad de la variable Y.

Planteemos el cálculo de otro tipo de distribución condicionada. Por ejemplo, estudiemos la distribución de los ingresos, X, para las familias cuyos gastos en alimentación sean inferiores a la mitad de sus ingresos. Es decir, se trata de hallar la distribución de X condicionada al suceso $\left[Y < \frac{X}{2}\right]$. En este caso, sin más que tener en cuenta las probabilidades condicionadas de sucesos, podemos hallar la función de distribución condicionada. Así,

$$F_{(X \mid X>2Y)}(x) = p[X \leq x \mid X > 2Y] = \frac{p[X \leq x, X > 2Y]}{p[X > 2Y]} .$$

En la figura 5.16 visualizamos el suceso $[X \leq x, X > 2Y]$ (sombreado oscuro) y el suceso $[X > 2Y]$ (líneas oblicuas).

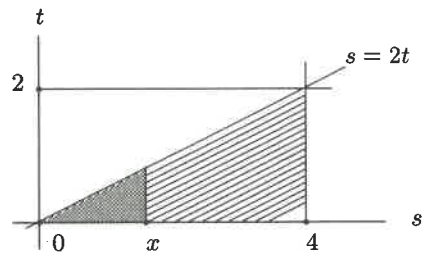


FIG. 5.16.

Calculemos las probabilidades de los mismos:

$$p[X > 2Y] = \int_{x=0}^{4} \int_{y=0}^{x/2} \frac{1}{24}(x+y)\, dydx = \frac{1}{24} \int_{x=0}^{4} \frac{5x^2}{8} dx = \frac{5}{9}$$

y, para los valores de x comprendidos entre 0 y 4,

$$p[X \leq x, X > 2Y] = \int_{s=0}^{x} \int_{t=0}^{s/2} \frac{1}{24}(s+t)\, dtds = \frac{5x^3}{576} .$$

Por tanto,

$$F_{(X \mid X>2Y)}(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^3}{64} & 0 < x < 4 \\ 1 & x \geq 4 \end{cases}$$

y su derivada

$$f_{(X \mid X>2Y)}(x) = \begin{cases} \frac{3x^2}{64} & 0 < x < 4 \\ 0 & \text{en el resto} \end{cases}$$

es la función de densidad. ◁

Veamos un ejemplo en el que las distribuciones condicionadas y marginales coinciden.

Ejemplo 5.15. Consideremos la variable (X, Y) estudiada en los ejemplos 5.7 y 5.10 cuya función de densidad es:

$$f(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Como ya vimos, la función de densidad marginal de Y es:

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Por tanto, la distribución de X condicionada a $Y = y$ tendrá la siguiente función de densidad:

$$f_{(X \mid Y=y)}(x) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1} = 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

siempre y cuando $f_Y(y) \neq 0$, es decir, para valores de la variable Y entre 0 y 1.

Análogamente se calcularía $f_{(Y \mid X=x)}(y)$.

Advierta el lector que, tanto para X como para Y, las distribuciones marginales coinciden con las condicionadas; es decir, la información completa sobre el valor que toma una de las variables, no altera la distribución de probabilidad de la otra.

Si estamos interesados en la obtención de funciones de densidad condicionadas por sucesos de probabilidad distinta de cero, tal es el caso, por ejemplo, de $f_{(X \mid Y<\frac{1}{2})}(x)$, hemos de hallar primero la correspondiente función de distribución condicionada, es decir,

$$F_{(X \mid Y<\frac{1}{2})}(x) = p\left[X \leq x \mid Y < \frac{1}{2}\right] = \frac{p[X \leq x, Y < \frac{1}{2}]}{p[Y < \frac{1}{2}]} .$$

Por un lado,

$$p\left[Y < \frac{1}{2}\right] = \int_0^{\frac{1}{2}} 1 \, dy = \frac{1}{2} .$$

Por otro lado, dado que la función de distribución conjunta es:

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \quad \text{o} \quad y < 0 \\ x \cdot y & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ x & \text{si } 0 \leq x \leq 1, y > 1 \\ y & \text{si } x > 1, 0 \leq y \leq 1 \\ 1 & \text{si } x > 1, y > 1 \end{cases}$$

obtenemos

$$p\left[X \leq x, Y < \frac{1}{2}\right] = F\left(x, \frac{1}{2}\right) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x \leq 1 \\ \frac{1}{2} & x > 1 \end{cases}$$

Por tanto,

$$F_{(X \mid Y<\frac{1}{2})}(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Derivando respecto de x, obtendremos la correspondiente función de densidad

$$f_{(X \mid Y<\frac{1}{2})}(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

Obsérvese que, de nuevo, la distribución de X condicionada a Y, en este caso $\left[Y \leq \frac{1}{2}\right]$, coincide con la distribución marginal de X. ◁

Todos los resultados anteriores pueden generalizarse para una variable aleatoria n-dimensional, $(X_1, \ldots, X_n)$. Téngase en cuenta que, en este caso, es posible plantear la distribución de un grupo de variables, $(X_{i_1}, \ldots, X_{i_k})$, condicionada a que la variable n-dimensional pertenezca a un cierto subconjunto de Borel de $\mathbb{R}^n$.

5.6 Independencia de variables aleatorias

Parece razonable pensar que existen variables que no se influyen en su comportamiento probabilístico; así, la variable X, *número de accidentes diarios en una cierta autopista*, no tiene por qué verse influida por la variable Y, *consumo mensual de una familia*. En cambio, existen otras variables en las que resulta lógico suponer un cierto grado de dependencia en su comportamiento; de este modo, la citada variable Y sí debería estar influida por la variable Z, *renta familiar*. A lo largo de este apartado, trataremos de formalizar esta idea intuitiva de influencia, o no, en el comportamiento aleatorio de las variables, es decir, de dependencia o independencia entre las mismas.

El concepto de independencia de variables aleatorias nos remite al de independencia de sucesos. En el capítulo 3 dijimos que dos sucesos, A y B, de un cierto espacio probabilístico, (Ω, α, p), eran independientes cuando la probabilidad de que ocurrieran simultáneamente coincidía con el producto de sus probabilidades individuales, esto es, cuando

$$p(A \cap B) = p(A) \cdot p(B) \, .$$

Supongamos que disponemos de una variable aleatoria bidimensional, (X, Y), sobre un espacio probabilístico (Ω, α, p), y sean ahora A y B dos subconjuntos de números reales de la σ-álgebra de Borel, β. Consideremos los sucesos $[X \in A]$ e $[Y \in B]$. Parece razonable relacionar la independencia de las variables X e Y con la de los sucesos $[X \in A]$ e $[Y \in B]$ dentro del espacio probabilístico (Ω, α, p). En definitiva, puede concebirse la independencia de las dos variables aleatorias como la independencia entre los sucesos referentes a X y los referentes a Y. Ello implicaría garantizar que se cumple la igualdad

$$p([X \in A] \cap [Y \in B]) = p[X \in A] \cdot p[Y \in B] \, .$$

Como la intersección de ambos sucesos es $[X \in A, Y \in B]$, la igualdad anterior puede escribirse como

$$p[X \in A, Y \in B] = p[X \in A] \cdot p[Y \in B] .$$

Traslademos esta idea a la siguiente definición.

Definición 5.16 *Sea (X, Y) una variable aleatoria bidimensional definida sobre el espacio probabilístico (Ω, α, p). Decimos que las variables aleatorias X e Y son independientes si, para cualesquiera subconjuntos A y B de la σ-álgebra de Borel, β, los sucesos $[X \in A]$ e $[Y \in B]$ son independientes; esto es, si*

$$p[X \in A, Y \in B] = p[X \in A] \cdot p[Y \in B] .$$

Diremos que las variables X e Y son dependientes si no verifican la definición anterior.

Esta definición resulta intuitiva ya que remite, como antes decíamos, la independencia de las variables aleatorias a la de los sucesos que generan; no obstante, resulta difícil de comprobar en las situaciones prácticas. El siguiente resultado, que sólo demostraremos parcialmente, indica que la independencia de dos variables aleatorias puede estudiarse a partir de su función de distribución conjunta.

Proposición 5.11 *Sea (X, Y) una variable aleatoria bidimensional y sea F su función de distribución conjunta. Entonces, X e Y son independientes si y sólo si, para todo punto (x, y) del plano, se cumple que*

$$F(x, y) = F_X(x) \cdot F_Y(y) ,$$

donde F_X y F_Y son las funciones de distribución marginales de X e Y, respectivamente.

Demostración. Probaremos únicamente que, si X e Y son independientes, la función de distribución conjunta se puede factorizar como producto de las marginales. La demostración general del resultado recíproco, esto es, que, si la función de distribución conjunta es igual al producto de las marginales, entonces, las variables aleatorias son independientes, desborda el nivel de este libro. Con todo, más adelante probaremos que el resultado es cierto para variables aleatorias discretas o continuas.

Supongamos que X e Y son independientes. Entonces, si A y B pertenecen a β,

$$p[X \in A, Y \in B] = p[X \in A] \cdot p[Y \in B] .$$

Sea (x, y) un punto cualquiera del plano. Si consideramos $A = (-\infty, x]$ y $B = (-\infty, y]$, y aplicamos la igualdad anterior,

$$p[X \leq x, Y \leq y] = p[X \leq x] \cdot p[Y \leq y] ,$$

esto es,

$$F(x, y) = F_X(x) \cdot F_Y(y) .$$

En definitiva, si X e Y son independientes, la función de distribución conjunta se puede factorizar como producto de las marginales, lo que prueba una de las implicaciones de la proposición. ♣

Como antes dijimos, se demuestra, también, la proposición recíproca, es decir, si la función de distribución conjunta se puede descomponer como producto de las marginales, las variables X e Y son independientes. El siguiente ejemplo ilustra parcialmente este resultado que, como ya hemos apuntado, probaremos más adelante para variables bidimensionales discretas o continuas.

Ejemplo 5.16. Sean X e Y dos variables aleatorias tales que, para todo punto, (x, y), se cumple:

$$F(x, y) = F_X(x) \cdot F_Y(y) \ .$$

Veamos que dichas variables son independientes. Para ello, consideremos x_1, x_2, y_1 e y_2 cuatro números reales cualesquiera con $x_1 < x_2$ e $y_1 < y_2$. Vamos a demostrar que, entonces,

$$p[x_1 < X \leq x_2, y_1 < Y \leq y_2] = p[x_1 < X \leq x_2] \cdot p[y_1 < Y \leq y_2] \ ,$$

esto es, que los sucesos $[x_1 < X \leq x_2]$ e $[y_1 < Y \leq y_2]$ son independientes. Adviértase que la independencia exigiría probar el resultado para dos subconjuntos de Borel cualesquiera, y no sólo para intervalos, si bien, dada la dificultad de dicha demostración, nosotros sólo la realizaremos para el citado caso particular.

En efecto, como vimos al estudiar las propiedades de la función de distribución conjunta,

$$p[x_1 < X \leq x_2, y_1 < Y \leq y_2] = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \ .$$

Pero, si F se descompone como producto de las marginales,

$$\begin{aligned} &p[x_1 < X \leq x_2, y_1 < Y \leq y_2] = \\ &\quad = F_X(x_2) \cdot F_Y(y_2) - F_X(x_1) \cdot F_Y(y_2) - F_X(x_2) \cdot F_Y(y_1) + F_X(x_1) \cdot F_Y(y_1) = \\ &\quad = F_X(x_2) \cdot [F_Y(y_2) - F_Y(y_1)] - F_X(x_1) \cdot [F_Y(y_2) - F_Y(y_1)] = \\ &\quad = [F_X(x_2) - F_X(x_1)] \cdot [F_Y(y_2) - F_Y(y_1)] = \\ &\quad = p[x_1 < X \leq x_2] \cdot p[y_1 < Y \leq y_2] \ , \end{aligned}$$

con lo cual queda probado que dos intervalos cualesquiera de valores de X e Y son independientes.

Por extensión a todo tipo de sucesos de Borel, deduciríamos la independencia de las variables X e Y. $\triangleleft$

La proposición 5.11 revela el hecho importante de que, si dos variables aleatorias son independientes, su función de distribución conjunta puede construirse con el conocimiento, únicamente, de las distribuciones marginales de cada una de las variables.

No obstante, la caracterización de la independencia de variables aleatorias mediante la función de distribución tiene poca relevancia práctica, ya que raramente se utiliza la función de distribución de una variable aleatoria bidimensional. Para variables discretas o continuas, dicha caracterización se realiza a través de la función de probabilidad o de la función de densidad, respectivamente, como muestran las dos proposiciones siguientes.

Proposición 5.12 *Sea (X, Y) una variable aleatoria bidimensional discreta y sea f su función de probabilidad. Entonces, X e Y son independientes si y sólo si, para todo valor (x_i, y_j) de (X, Y), se cumple que*

$$f(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j),$$

donde f_X y f_Y son las funciones de probabilidad de X e Y, respectivamente.

Demostración. Supongamos que X e Y son independientes. Sabemos entonces que, si A y B pertenecen a β,

$$p[X \in A, Y \in B] = p[X \in A] \cdot p[Y \in B] .$$

Entonces, si tomamos $A = \{x_i\}$ y $B = \{y_j\}$, y concretamos la igualdad anterior,

$$f(x_i, y_j) = p[X = x_i, Y = y_j] = p[X = x_i] \cdot p[Y = y_j] = f_X(x_i) \cdot f_Y(y_j) ,$$

lo que demuestra una de las implicaciones.

Supongamos, ahora, que la función de probabilidad conjunta es igual al producto de las marginales. Entonces, dados dos subconjuntos A y B de Borel de $\mathbb{R}$,

$$\begin{aligned} p[X \in A, Y \in B] &= \textstyle\sum_{x_i \in A} \sum_{y_j \in B} f(x_i, y_j) = \sum_{x_i \in A} \sum_{y_j \in B} f_X(x_i) \cdot f_Y(y_j) = \\ &= \textstyle\sum_{x_i \in A} f_X(x_i) \cdot \sum_{y_j \in B} f_Y(y_j) = p[X \in A] \cdot p[Y \in B] . \end{aligned}$$

Por tanto, X e Y son independientes.

Obsérvese que la condición de independencia se puede enunciar para cualquier punto (x, y) del plano, pues si x o y no son valores de X o de Y, respectivamente, la factorización de la función de probabilidad es obvia, dado que se anulan, simultáneamente, los dos lados de la igualdad. ♣

Veamos, a continuación, la proposición análoga para variables bidimensionales continuas.

Proposición 5.13 *Sea (X, Y) una variable aleatoria bidimensional continua y sea f su función de densidad. Entonces, X e Y son independientes si y sólo si, para todo punto (x, y) del plano, se cumple que*

$$f(x, y) = f_X(x) \cdot f_Y(y),$$

donde f_X y f_Y son las funciones de densidad marginales de X e Y, respectivamente.

Demostración. Supongamos que X e Y son independientes. Por la proposición 5.11 se cumplirá que

$$F(x, y) = F_X(x) \cdot F_Y(y) ,$$

para todo punto (x, y) del plano. Si derivamos con respecto a ambas variables, como $f(x, y) = \partial^2 F(x, y)/\partial x \partial y$, entonces

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F_X(x) \cdot F_Y(y) = \frac{d}{dx} F_X(x) \cdot \frac{d}{dy} F_Y(y) = f_X(x) \cdot f_Y(y) .$$

Supongamos, ahora, que la función de densidad conjunta es igual al producto de las marginales. Entonces, dados dos subconjuntos, A y B, de Borel,

$$\begin{aligned} p[X \in A, Y \in B] &= \int_A \int_B f(x, y) dx dy = \int_A \int_B f_X(x) \cdot f_Y(y) dx dy = \\ &= \int_A f_X(x) dx \cdot \int_B f_Y(y) dy = p[X \in A] \cdot p[Y \in B] \end{aligned}$$

de lo que se concluye que X e Y son independientes. ♣

Las caracterizaciones de la independencia de variables aleatorias que acabamos de demostrar son las que se utilizan en la mayoría de las aplicaciones prácticas. Por un lado, cuando se conoce la función de probabilidad o la función de densidad conjunta, sirven para estudiar si dos variables aleatorias son, o no, independientes. Pero, además, cuando la independencia de las variables pueda aceptarse como hipótesis de trabajo, se utilizan para construir la función de probabilidad o de densidad conjunta, conocidas sólo las distribuciones marginales. Ello ocurre cuando, por la *naturaleza* de las variables dentro del problema concreto en estudio, pueda presuponerse que cada una de las variables no influye en la otra. En esas situaciones, dicho supuesto se convierte en una aceptación de la independencia estadística entre ambas variables.

En el siguiente ejemplo probaremos la independencia de dos variables discretas en virtud de la descomposición factorial de su función de probabilidad conjunta.

Ejemplo 5.17. Se realizan lanzamientos independientes de una moneda no cargada. Sea X la variable aleatoria que representa el *número de lanzamientos requeridos para obtener la primera cara*, e Y la que representa el *número de lanzamientos que hay que realizar posteriormente hasta obtener la segunda cara.* Por ejemplo, si en los cinco primeros lanzamientos se obtiene el resultado $zczzc$ (caras en el segundo y quinto lanzamiento, y cruces en los restantes), el valor de X sería 2 (las dos primeras tiradas) y el de Y sería 3 (las tres últimas).

X e Y son variables discretas que toman como valores todos los enteros positivos. Su función de probabilidad se obtiene fácilmente. Por ejemplo,

$$p[X=1,Y=1]=p(cc)=p(c)\cdot p(c)=\frac{1}{2^2}\,,$$

$$p[X=2,Y=1]=p(zcc)=p(z)\cdot p(c)\cdot p(c)=\frac{1}{2^3}$$

y, en general,

$$p[X=i,Y=j]=p\left(z\ldots{}^{(i-1)}\ldots zcz\ldots{}^{(j-1)}\ldots zc\right)=\frac{1}{2^{i+j}}\,.$$

La función de probabilidad conjunta se puede escribir en la siguiente tabla de doble entrada:

$X\backslash Y$	1	2	3	$\ldots$	
1	$1/2^2$	$1/2^3$	$1/2^4$	$\ldots$	$1/2$
2	$1/2^3$	$1/2^4$	$1/2^5$	$\ldots$	$1/2^2$
3	$1/2^4$	$1/2^5$	$1/2^6$	$\ldots$	$1/2^3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	$1/2$	$1/2^2$	$1/2^3$	$\ldots$	1

Sumando filas y sumando columnas en dicha tabla, se calculan, sin dificultad, las funciones de probabilidad marginales:

$$p[X=i]=\sum_{j=1}^{\infty}p[X=i,Y=j]=\sum_{j=1}^{\infty}\frac{1}{2^{i+j}}=\frac{1}{2^i}\,,\qquad i=1,2,\ldots$$

y, análogamente,

$$p[Y = j] = \frac{1}{2^j}\,, \qquad j = 1, 2, \ldots .$$

Por tanto, para todo (i, j),

$$p[X = i] \cdot p[Y = j] = \frac{1}{2^i} \cdot \frac{1}{2^j} = \frac{1}{2^{i+j}} = p[X = i, Y = j]\,.$$

Es decir, las variables son independientes. ◁

Con el siguiente ejemplo se ilustra, para variables continuas, la construcción de la función de densidad conjunta a partir de la suposición de independencia.

Ejemplo 5.18. Las propinas de un camarero en el mes de enero, en miles de pesetas, X, se supone que siguen una distribución continua con función de densidad

$$f_X(x) = \begin{cases} \frac{1}{50} & \text{si } 0 < x < 50 \\ 0 & \text{en el resto} \end{cases}$$

Las propinas durantes el mes de febrero (asimismo en miles), Y, se supone que siguen la misma distribución. Obtengamos la probabilidad de que las propinas en ambos meses sumen al menos 60.000 pesetas, suponiendo que ambas variables son independientes.

Si consideramos que X e Y son independientes (es decir, son variables aleatorias idénticamente distribuidas y, además, independientes), la función de densidad conjunta de X e Y será

$$f(x, y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{2500} & \text{si } 0 < x < 50,\ 0 < y < 50 \\ 0 & \text{en el resto} \end{cases}$$

En consecuencia, integrando esta función en el recinto en que $[X + Y \geq 60]$ (véase la figura 5.17), se obtiene

$$p[X + Y \geq 60] = \int_{10}^{50} \int_{60-x}^{50} \frac{1}{2500} dy dx = \frac{8}{25}\,.$$

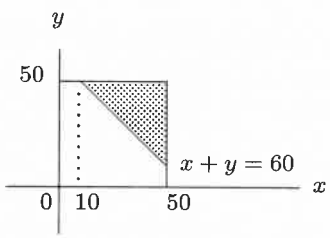


FIG. 5.17.

De la misma manera, el lector puede calcular otras probabilidades; por ejemplo, la de que las propinas de febrero no lleguen a triplicar las de enero, es decir, $p[Y < 3X]$. ◁

Como antes dijimos, la demostración de que la descomposición factorial de la función de distribución conjunta en sus funciones de distribución marginales, implica la independencia de dos variables aleatorias, rebasa el nivel de este libro. No obstante, resulta sencilla de probar para variables aleatorias bidimensionales discretas o continuas, según veremos a continuación.

Proposición 5.14 *Sea (X,Y) una variable aleatoria bidimensional discreta o continua y sea F su función de distribución conjunta. Entonces, X e Y son independientes si y sólo si, para todo punto (x,y) del plano, se cumple que*

$$F(x,y) = F_X(x) \cdot F_Y(y) \ ,$$

donde F_X y F_Y son las funciones de distribución marginales de X e Y, respectivamente.

Demostración. En la proposición 5.11 demostramos que, si X e Y son independientes, la función de distribución conjunta se descompone como producto de las marginales. Veamos, ahora, el resultado recíproco.

Supongamos que (X,Y) es una variable bidimensional discreta y que, para todo punto (x,y) del plano, se cumple:

$$F(x,y) = F_X(x) \cdot F_Y(y) \ .$$

Entonces, teniendo en cuenta la proposición 5.4,

$$\begin{aligned} f(x_i,y_j) &= p[X=x_i, Y=y_j] = \\ &= F(x_i,y_j) - F(x_i,y_{j-1}) - F(x_{i-1},y_j) + F(x_{i-1},y_{j-1}) \ , \end{aligned}$$

que, en virtud de la hipótesis, resulta

$$F_X(x_i) \cdot F_Y(y_j) - F_X(x_i) \cdot F_Y(y_{j-1}) - F_X(x_{i-1}) \cdot F_Y(y_j) + F_X(x_{i-1}) \cdot F_Y(y_{j-1}) \ ,$$

y, agrupando los términos correspondientes,

$$[F_X(x_i) - F_X(x_{i-1})] \cdot [F_Y(y_j) - F_Y(y_{j-1})] = p[X=x_i] \cdot p[Y=y_j] = f_X(x_i) \cdot f_Y(y_j) \ .$$

Si x_i o y_j fuesen los valores más pequeños de X y de Y, respectivamente, la demostración sería análoga.

Por consiguiente, la descomposición factorial de la función de distribución conjunta implica la de la función de probabilidad conjunta. Como esta última equivalía a la independencia (proposición 5.13), queda probado el resultado para variables discretas.

Consideremos, ahora, que (X,Y) es continua. Puesto que en la proposición 5.12 demostrábamos que la descomposición de la función de distribución implicaba la de la función de densidad conjunta y que ésta implicaba, a su vez, la independencia, la demostración de este caso está incluida en la de dicha proposición. ♣

Las tres últimas proposiciones se resumen en el siguiente resultado que, aun no siendo general, resulta suficiente para casi todas las aplicaciones.

Proposición 5.15 *Sea (X,Y) una variable bidimensional discreta o continua. Entonces, son equivalentes las siguientes afirmaciones:*

1. *X e Y son independientes.*
2. *$F(x,y) = F_X(x) \cdot F_Y(y)$ para todo punto (x,y) del plano.*
3. *$f(x,y) = f_X(x) \cdot f_Y(y)$ para todo punto (x,y) del plano.*

En definitiva, para una variable aleatoria bidimensional discreta o continua, la comprobación de la independencia se realiza mediante la función de probabilidad, o de densidad, conjunta. Una vez obtenidas a partir de ella las marginales, se comprueba si su producto coincide o no con la conjunta. A continuación, trataremos de buscar otras formas más sencillas de caracterizar la independencia entre variables.

Abordemos, en primer lugar, el caso en que ambas variables sean discretas. Como puede observarse en el ejemplo 5.17, las filas de la tabla de doble entrada son proporcionales. Así, si multiplicamos la primera fila por $1/2$ obtenemos la segunda, si a ésta la multiplicamos por $1/2$ obtenemos la tercera, y así sucesivamente. Otro tanto ocurre para las columnas. Seguidamente comprobaremos que este hecho no es casual.

Según demostraremos, si las variables son independientes, el cociente entre los elementos de las filas i-ésima y k-ésima de la tabla de doble entrada es el mismo para cada columna j-ésima; las filas de la tabla serían, en consecuencia, proporcionales. Otro tanto ocurre para las columnas. Veremos que también se cumple el resultado recíproco; esto es, si las filas o las columnas de la tabla de doble entrada son proporcionales, las variables son independientes.

Por tanto, para una variable bidimensional discreta, la proporcionalidad entre filas (o entre columnas) de la tabla de doble entrada es una forma sencilla de comprobar su independencia.

La siguiente proposición prueba este resultado.

Proposición 5.16 *Sea (X,Y) una variable aleatoria bidimensional discreta. Entonces, X e Y son independientes si y sólo si, para cualquier par de valores x_i y x_k de X, se verifica que*

$$\frac{p[X = x_i, Y = y_j]}{p[X = x_k, Y = y_j]}$$

es constante para todo valor y_j de Y.

Demostración. Si X e Y son independientes, para cualesquiera x_i y x_k de X,

$$\frac{p[X = x_i, Y = y_j]}{p[X = x_k, Y = y_j]} = \frac{p[X = x_i] \cdot p[Y = y_j]}{p[X = x_k] \cdot p[Y = y_j]} = \frac{p[X = x_i]}{p[X = x_k]},$$

por tanto, dicho cociente es constante sea cual sea y_j de Y.

Si, para cualquier par de valores, x_i y x_k, de X,

$$\frac{p[X = x_i, Y = y_j]}{p[X = x_k, Y = y_j]}$$

es constante para todo y_j de Y, entonces[4],

$$\frac{p[X=x_i,Y=y_1]}{p[X=x_k,Y=y_1]}=\cdots=\frac{p[X=x_i,Y=y_m]}{p[X=x_k,Y=y_m]}=\cdots=\frac{\sum_{j=1}^{\infty}p[X=x_i,Y=y_j]}{\sum_{j=1}^{\infty}p[X=x_k,Y=y_j]},$$

es decir,

$$\frac{p[X=x_i,Y=y_j]}{p[X=x_k,Y=y_j]}=\frac{p[X=x_i]}{p[X=x_k]}$$

para todo y_j. Por tanto,

$$\frac{p[X=x_i,Y=y_j]}{p[X=x_i]}=\frac{p[X=x_k,Y=y_j]}{p[X=x_k]}.$$

Aplicando de nuevo el resultado de la nota 4 a pie de página,

$$\frac{p[X=x_1,Y=y_j]}{p[X=x_1]}=\cdots=\frac{p[X=x_m,Y=y_j]}{p[X=x_m]}=\cdots=\frac{\sum_{i=1}^{\infty}p[X=x_i,Y=y_j]}{\sum_{i=1}^{\infty}p[X=x_i]},$$

es decir,

$$\frac{p[X=x_i,Y=y_j]}{p[X=x_i]}=p[Y=y_j].$$

De aquí,

$$p[X=x_i,Y=y_j]=p[X=x_i]\cdot p[Y=y_j],$$

para cualquier par de valores x_i e y_j.

Por tanto, X e Y son independientes. ♣

Igualmente se podría enunciar la proposición en términos de la proporcionalidad de las columnas; es decir, *X e Y son independientes si y sólo si, para cualquier par de valores y_j e y_l de Y, se verifica que*

$$\frac{p[X=x_i,Y=y_j]}{p[X=x_i,Y=y_l]}$$

es constante para todo valor x_i de X.

Para variables bidimensionales continuas existe también un procedimiento sencillo de comprobar la independencia de dos variables aleatorias sin necesidad de calcular las funciones de densidad marginales. Este procedimiento se detalla en la siguiente proposición.

4. Un resultado clásico referente a proporciones indica que, si

$$\frac{a_1}{b_1}=\cdots=\frac{a_n}{b_n}=\cdots,$$

entonces esos cocientes son, asimismo, iguales a

$$\frac{\sum a_n}{\sum b_n}.$$

Proposición 5.17 *Sea (X,Y) una variable aleatoria bidimensional continua y sea f su función de densidad conjunta. Entonces, X e Y son independientes si y sólo si, para todo punto (x,y) del plano,*

$$f(x,y) = g_1(x) \cdot g_2(y) \ ,$$

donde g_1 es una función que no depende de y, y g_2 una función que no depende de x.

Demostración. Obviamente, si X e Y son independientes la función de densidad conjunta admite la descomposición del enunciado, sin más que tomar f_X como g_1 y f_Y como g_2. Veamos el resultado recíproco.

Supongamos que la función de densidad conjunta admite la descomposición del enunciado. Obtengamos las funciones de densidad marginales:

$$f_X(x) = \int_{-\infty}^{\infty} g_1(x) \cdot g_2(y)dy = g_1(x) \cdot \int_{-\infty}^{\infty} g_2(y)dy$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} g_1(x) \cdot g_2(y)dx = g_2(y) \cdot \int_{-\infty}^{\infty} g_1(x)dx \ .$$

Ahora bien,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) \cdot g_2(y)dxdy = \int_{-\infty}^{\infty} g_1(x)dx \cdot \int_{-\infty}^{\infty} g_2(y)dy \ .$$

Pero esta integral es igual a la unidad y, por tanto,

$$\begin{aligned} f_X(x) \cdot f_Y(y) &= \left[g_1(x) \cdot \int_{-\infty}^{\infty} g_2(y)dy\right] \cdot \left[g_2(y) \cdot \int_{-\infty}^{\infty} g_1(x)dx\right] = \\ &= g_1(x) \cdot g_2(y) \cdot \int_{-\infty}^{\infty} g_2(y)dy \cdot \int_{-\infty}^{\infty} g_1(x)dx = f(x,y) \end{aligned}$$

y, en consecuencia, X e Y son independientes. ♣

Veamos el siguiente ejemplo.

Ejemplo 5.19. Sea (X,Y) una variable bidimensional continua con función de densidad conjunta

$$f(x,y) = \begin{cases} \frac{9}{4}x^2 \cdot y^2 & -1 \le x \le 1, \quad -1 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

Si tomamos

$$g_1(x) = \begin{cases} \frac{9}{4}x^2 & -1 \le x \le 1 \\ 0 & \text{en el resto} \end{cases}$$

y

$$g_2(y) = \begin{cases} y^2 & -1 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

se cumplirá, para todo punto (x,y) del plano, la igualdad

$$f(x,y) = g_1(x) \cdot g_2(y) \ .$$

Las variables X e Y son, por consiguiente, independientes en virtud de la proposición anterior.

Puede comprobarse este hecho calculando las funciones de densidad marginales. Integrando la función de densidad conjunta,

$$f_X(x) = \int_{-1}^{1} \frac{9}{4}x^2 \cdot y^2 dy = \begin{cases} \frac{3}{2}x^2 & -1 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \int_{-1}^{1} \frac{9}{4}x^2 \cdot y^2 dx = \begin{cases} \frac{3}{2}y^2 & -1 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Se cumple, así, la descomposición de la función de densidad conjunta como producto de las marginales. ◁

Conviene, con todo, precisar que la descomposición factorial de la proposición 5.17 exige que se pueda realizar en todo el plano; ello puede crear problemas cuando el campo de variación de la variable bidimensional (esto es, el subconjunto del plano en el que la función de densidad es no nula) depende de x y de y, sin que se puedan *separar* ambas dependencias. Obsérvese lo que ocurre en el siguiente ejemplo que, aparentemente, es similar al anterior.

Ejemplo 5.20. Sea (X, Y) una variable bidimensional continua con función de densidad conjunta

$$f(x, y) = \begin{cases} 2x \cdot y & \text{si } (x, y) \in S \\ 0 & \text{en el resto} \end{cases}$$

siendo S el recinto que se describe en la figura 5.18.

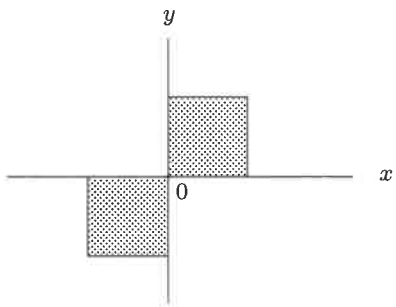


FIG. 5.18.

Aunque el aspecto funcional de esta densidad conjunta es parecido al del ejemplo anterior, y sugiere una descomposición similar (por ejemplo, $g_1(x) = 2x$ y $g_2(y) = y$), esta descomposición no puede realizarse en todo el plano porque el campo de variación de la variable bidimensional (un cuadrado en el primer cuadrante y otro en el tercero) no puede separarse para cada una de las variables (no es un rectángulo del plano). En concreto, si $x = -0.5$, y varía entre -1 y 0; en cambio, si $x = 0.5$, dicha variación es entre 0 y 1.

Si se integra la función de densidad conjunta, se obtienen sin dificultad las marginales:

$$f_X(x) = \begin{cases} -x & -1 \leq x < 0 \\ x & 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \begin{cases} -y & -1 \leq y < 0 \\ y & 0 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

es decir,

$$f_X(x) = \begin{cases} |x| & -1 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \begin{cases} |y| & -1 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

En consecuencia, las variables no son independientes, ya que el producto de las funciones de densidad marginales vale

$$f_X(x) \cdot f_Y(y) = \begin{cases} |x| \cdot |y| & -1 \leq x \leq 1,\ -1 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

que no coincide con la función de densidad conjunta. ◁

Como es razonable esperar desde el punto de vista intuitivo, la independencia de dos variables aleatorias se transmite a sus funciones. Demostremos el siguiente resultado que confirma ese hecho.

Proposición 5.18 *Sea (X,Y) una variable aleatoria bidimensional con X e Y variables independientes, y sean $U = g(X)$ y $V = h(Y)$ con g y h funciones reales tales que U y V sean variables aleatorias. Entonces, U y V son independientes.*

Demostración. Sean A y B dos subconjuntos de Borel. El suceso $[U \in A]$ puede escribirse como $[g(X) \in A]$ y, por tanto, coincide con el suceso $[X \in g^{-1}(A)]$. Análogamente, podemos escribir el suceso $[V \in B]$ como $[Y \in h^{-1}(B)]$. En consecuencia,

$$p[U \in A, V \in B] = p[X \in g^{-1}(A), Y \in h^{-1}(B)] .$$

Como X e Y son independientes,

$$p[X \in g^{-1}(A), Y \in h^{-1}(B)] = p[X \in g^{-1}(A)] \cdot p[Y \in h^{-1}(B)] = p[U \in A] \cdot p[V \in B] .$$

En definitiva,

$$p[U \in A, V \in B] = p[U \in A] \cdot p[V \in B] ,$$

por lo que U y V son independientes. ♣

Anteriormente indicamos que la independencia de dos variables, X e Y, podía interpretarse desde el punto de vista intuitivo como el hecho de que cada una de las variables no influía en la otra. Parece entonces razonable, continuando con esa interpretación, que el conocimiento del valor de la variable Y, por ejemplo, no modifique las probabilidades de sucesos referentes a la variable X. En definitiva, y como

muestran los siguientes resultados, la distribución de cada una de las variables condicionada por la otra no va a depender del suceso que condicione la distribución y va a coincidir, por tanto, con la distribución marginal. En el fondo, esta característica va a ser una definición equivalente a la de independencia de variables aleatorias.

Proposición 5.19 *Sea (X,Y) una variable aleatoria bidimensional discreta. Entonces, X e Y son independientes si y sólo si*

$$f_{(X \mid Y=y)}(x) = f_X(x) ,$$

para todo número real x y para cualquier valor y de Y.

Demostración. Recordemos que, si y es un valor de la variable aleatoria Y, entonces $p[Y = y] \neq 0$ y, por tanto, la función de probabilidad condicionada del enunciado está definida.

Supongamos que X e Y son independientes. Entonces,

$$f_{(X \mid Y=y)}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x) .$$

Recíprocamente, admitamos que se cumple la igualdad del enunciado. En ese caso, para todo número real x,

$$f(x,y) = f_{(X \mid Y=y)}(x) \cdot f_Y(y) = f_X(x) \cdot f_Y(y) ,$$

es decir, X e Y son independientes. ♣

Análogamente, se demuestra que *X e Y son independientes si y sólo si*

$$f_{(Y \mid X=x)}(y) = f_Y(y) ,$$

para todo número real y y para cualquier valor x de X.

La siguiente proposición prueba un resultado similar para variables continuas que se establece, ahora, en términos de la función de densidad.

Proposición 5.20 *Sea (X,Y) una variable aleatoria bidimensional continua. Entonces, X e Y son independientes si y sólo si*

$$f_{(X \mid Y=y)}(x) = f_X(x) ,$$

para todo número real x y para cualquier valor y de Y.

Demostración. Si X e Y son independientes, para todo punto (x,y) del plano,

$$f(x,y) = f_X(x) \cdot f_Y(y) .$$

En consecuencia, si y es un valor posible de Y (y, por tanto, $f_Y(y) \neq 0$),

$$f_{(X \mid Y=y)}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x) ,$$

para todo número real x.

Recíprocamente, supongamos que se cumple la igualdad del enunciado, y sea (x, y) un punto del plano. Entonces, si y es un valor de Y,

$$f(x,y) = f_{(X \mid Y=y)}(x) \cdot f_Y(y) = f_X(x) \cdot f_Y(y) \ .$$

Si y no es un valor de Y, $f_Y(y)$ será igual a cero, y también lo será $f(x,y)$, puesto que, en tal caso, (x,y) no será un valor de (X,Y) (de hecho, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = 0$ obliga a que $f(x,y)$ también se anule). En consecuencia, se cumple:

$$f(x,y) = f_X(x) \cdot f_Y(y) \ ,$$

porque ambos miembros de la igualdad son nulos. Por tanto, para cualquier punto del plano se verifica la descomposición de la función de densidad conjunta como producto de las marginales y, en definitiva, X e Y son independientes. ♣

Del mismo modo se demuestra que X *e* Y *son independientes si y sólo si*

$$f_{(Y \mid X=x)}(y) = f_Y(y) \ ,$$

para todo número real y *y para cualquier valor* x *de* X.

Las dos proposiciones anteriores prueban que, si dos variables aleatorias son independientes, la distribución de una de ellas condicionada por un valor concreto de la otra no depende de dicho condicionamiento y coincide, por tanto, con la marginal. Se demuestran resultados similares cuando los sucesos que condicionan son más amplios, esto es, de la forma $[Y \in B]$ o $[X \in A]$.

En los dos ejemplos siguientes se ilustra, para variables discretas y para continuas, la caracterización de la independencia a partir de las distribuciones condicionadas.

Ejemplo 5.21. Se realizan lanzamientos independientes de una moneda normal. Consideremos, de nuevo, las variables X, *número de lanzamientos requeridos para obtener la primera cara*, e Y, *número de lanzamientos que hay que realizar posteriormente hasta obtener la segunda cara*, cuya función de probabilidad conjunta, que se dedujo en el ejemplo 5.17, es

$$f(i,j) = p[X = i, Y = j] = \frac{1}{2^{i+j}} \quad \text{para todo } (i,j).$$

La función de probabilidad de X condicionada a $Y = j$ es

$$f_{(X \mid Y=j)}(i) = \frac{f(i,j)}{f_Y(j)} = \frac{\frac{1}{2^{i+j}}}{\frac{1}{2^j}} = \frac{1}{2^i} = p[X = i] = f_X(i) \quad i = 1, 2, \ldots$$

De acuerdo con esto, la distribución de X condicionada a cualquier valor de Y coincide con la marginal de X, dado que X e Y son, como ya vimos, independientes.

Si condicionamos X a cualquier suceso que sólo dependa de Y, el resultado va a ser idéntico; es decir, la distribución condicionada coincidiría con la marginal. En este sentido, el lector puede hallar la distribución de X condicionada a que Y tome un valor par y comprobar que ésta coincide con la distribución marginal de X. Ahora bien, si el suceso por el que se condiciona depende de ambas variables, lógicamente no se verificará lo anterior. Para comprobarlo, obtengamos la distribución de X condicionada a que X e Y tomen el mismo valor, es decir, $f_{(X \mid X=Y)}(x)$.

Calculemos, previamente, la probabilidad del suceso que condiciona:

$$\begin{aligned} p[X=Y] &= p[X=1,Y=1]+p[X=2,Y=2]+\cdots = \\ &= \sum_{i=1}^{\infty} p[X=i,Y=i] = \sum_{i=1}^{\infty} \frac{1}{2^{i+i}} = \frac{1}{3}\,, \end{aligned}$$

que no es otra cosa que la suma de los valores de la diagonal de la tabla de doble entrada que recoge la función de probabilidad de (X,Y). Entonces,

$$\begin{aligned} f_{(X \mid X=Y)}(x) = p[X=i \mid X=Y] &= \frac{p[X=i,X=Y]}{p[X=Y]} = \frac{p[X=i,Y=i]}{p[X=Y]} = \\ &= \frac{\frac{1}{2^{i+i}}}{\frac{1}{3}} = \frac{3}{2^{2i}} \quad i=1,2,\ldots, \end{aligned}$$

distribución que no coincide con la marginal de X. ◁

Ejemplo 5.22. Consideremos, de nuevo, la variable aleatoria del ejemplo 5.20, cuya función de densidad conjunta es

$$f(x,y) = \begin{cases} 2x\cdot y & -1 \le x < 0, \quad -1 \le y < 0 \\ & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

La función de densidad de Y condicionada a $X=-0.5$ es

$$f_{(Y \mid X=-0.5)}(y) = \frac{f(-0.5,y)}{f_X(-0.5)} = \frac{2\cdot(-0.5)\cdot y}{0.5} = \begin{cases} -2y & -1 \le y < 0 \\ 0 & \text{en el resto} \end{cases}$$

dado que

$$f_X(x) = \begin{cases} |x| & -1 \le x \le 1 \\ 0 & \text{en el resto} \end{cases}$$

Evidentemente, $f_{(Y \mid X=-0.5)}(y)$ no coincide con la marginal de Y calculada en el ejemplo 5.20; es decir, como ya sabíamos, X e Y son variables dependientes.

Las funciones de densidad de Y condicionadas a $X=x$, tantas como valores posibles de X, se calcularán de idéntica manera, resultando ser:

- Para un valor de x tal que $-1 \le x < 0$,

$$f_{(Y \mid X=x)}(y) = \frac{f(x,y)}{f_X(x)} = \begin{cases} -2y & -1 \le y < 0 \\ 0 & \text{en el resto} \end{cases}$$

- Para un valor de x tal que $0 < x \le 1$,

$$f_{(Y \mid X=x)}(y) = \frac{f(x,y)}{f_X(x)} = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{en el resto} \end{cases}$$

Comprobemos, a continuación, que la distribución de Y condicionada a valores positivos de X no coincide con la distribución de Y; es decir que

$$f_{(Y \mid X \ge 0)}(y) \neq f_Y(y)\,.$$

Recordemos que para hallar $f_{(Y \mid X \geq 0)}(y)$ hay que obtener previamente la función de distribución condicionada.

$$F_{(Y \mid X \geq 0)}(y) = p[Y \leq y \mid X \geq 0] = \frac{p[Y \leq y, X \geq 0]}{p[X \geq 0]} .$$

Para un valor de y tal que $0 \leq y \leq 1$, el numerador vale

$$p[Y \leq y, X \geq 0] = \int_0^1 \int_0^y 2st \, dt \, ds = \frac{y^2}{2} .$$

Por otra parte, el denominador, que se obtiene integrando $f_X(x)$ en el recinto donde $x \geq 0$, vale $1/2$.

En consecuencia,

$$F_{(Y \mid X \geq 0)}(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

Derivando esta función, resulta:

$$f_{(Y \mid X \geq 0)}(y) = \begin{cases} 2y & 0 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

función de densidad distinta de $f_Y(y)$. ◁

El conocimiento de las distribuciones de X y de Y ($f_X(x)$ y $f_Y(y)$, en términos de funciones de densidad o de probabilidad) no es, en principio, suficiente para determinar la distribución conjunta de las mismas ($f_{(X,Y)}(x, y)$, en los mismos términos). Para determinar su comportamiento aleatorio conjunto debemos de tener, además, información acerca de la relación probabilística entre X e Y. Esta relación puede venir dada conociendo:

- Directamente la distribución conjunta, $f_{(X,Y)}(x, y)$.
- Una distribución marginal (por ejemplo $f_X(x)$) y la distribución condicionada complementaria (en este caso, $f_{(Y \mid X=x)}(y)$), pues

$$f_X(x) \cdot f_{(Y \mid X=x)}(y) = f_{(X,Y)}(x, y) .$$

- Las dos distribuciones marginales y suponiendo una relación de independencia entre las variables. Entonces,

$$f_X(x) \cdot f_Y(y) = f_{(X,Y)}(x, y) .$$

Por tanto, el conocimiento único de las distribuciones marginales no determina la conjunta. Dicho de otro modo, dos distribuciones marginales determinadas pueden provenir de un sinfín de distribuciones conjuntas. Las definiciones y resultados anteriores se generalizan para variables aleatorias n-dimensionales. Veamos cómo.

Definición 5.17 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional definida sobre el espacio probabilístico (Ω, α, p). Decimos que $X_1, \ldots, X_n$ son independientes si, para cualesquiera subconjuntos $B_1, \ldots, B_n$ de la σ-álgebra de Borel, β, los n sucesos $[X_1 \in B_1], \ldots, [X_n \in B_n]$ son independientes.*

Como en el caso bidimensional, no es fácil comprobar que n variables son independientes utilizando esta definición. Los siguientes resultados, cuyas demostraciones son análogas a las realizadas en el caso bidimensional, indican que puede estudiarse la independencia de n variables a partir de su función de distribución y, si $(X_1, \ldots, X_n)$ es discreta o continua, a partir de su función de probabilidad o de densidad conjunta, respectivamente.

Proposición 5.21 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional y sea F su función de distribución conjunta. Entonces, $X_1, \ldots, X_n$ son independientes si y sólo si, para todo punto $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$, se cumple que*

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) ,$$

donde F_{X_i} es la función de distribución marginal de X_i.

Parece obvio que, si las variables $X_1, \ldots, X_n$ son independientes, también lo son cualquier grupo de ellas. En efecto, dado que se cumple que

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) ,$$

haciendo tender a $+\infty$ las variables que no entran en el grupo, se obtiene la factorización deseada. Por ejemplo, si queremos ver que son independientes las tres primeras,

$$F(x_1, x_2, x_3, +\infty, \ldots, +\infty) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3) \cdot F_{X_4}(+\infty) \cdots F_{x_n}(+\infty) ,$$

es decir,

$$F_{(X_1, X_2, X_3)}(x_1, x_2, x_3) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot F_{X_3}(x_3) .$$

Por tanto, X_1, X_2 y X_3 son variables independientes.

Por tanto, dadas n variables independientes (mutuamente independientes), también serán independientes dos a dos, tres a tres, etc. En cambio, puede haber, por ejemplo, tres variables independientes dos a dos pero no mutuamente independientes (véase el ejercicio propuesto 5.27).

Para variables discretas, podemos caracterizar la independencia a partir de la función de probabilidad.

Proposición 5.22 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional discreta y sea f su función de probabilidad. Entonces, $X_1, \ldots, X_n$ son independientes si y sólo si, para todo valor $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$, se cumple que*

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n),$$

donde f_{X_i} es la función de probabilidad de X_i.

Veamos la siguiente proposición análoga para variables n-dimensionales continuas.

Proposición 5.23 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional continua y sea f su función de densidad. Entonces, $X_1, \ldots, X_n$ son independientes si y sólo si, para todo punto $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$, se cumple que*

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n),$$

donde f_{X_i} es la función de densidad marginal de X_i.

En definitiva, para variables aleatorias n-dimensionales discretas o continuas, la independencia de variables equivale a la descomposición, tanto de la función de distribución, como de la función de densidad o de probabilidad, en sus respectivas marginales.

Proposición 5.24 *Sea $(X_1, \ldots, X_n)$ una variable n-dimensional discreta o continua. Entonces, son equivalentes las siguientes afirmaciones:*

1. *$X_1, \ldots, X_n$ son independientes.*
2. *$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$ para todo punto $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$.*
3. *$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ para todo punto $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$.*

Para comprobar la independencia de n variables aleatorias continuas puede ensayarse la descomposición de la función de densidad conjunta como producto de funciones de cada una de las variables, sin necesidad de obtener las funciones de densidad marginal. Con todo, conviene precisar que esta descomposición factorial debe realizarse en todo $\mathbb{R}^n$, por lo que, cuando se utilice el resultado que enunciamos a continuación, hay que ser especialmente cuidadoso en el estudio del campo de definición de las variables.

Proposición 5.25 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional continua y sea f su función de densidad conjunta. Entonces, $X_1, \ldots, X_n$ son independientes si y sólo si, para todo punto $(x_1, \ldots, x_n)$ de $\mathbb{R}^n$,*

$$f(x_1, \ldots, x_n) = g_1(x_1) \cdots g_n(x_n)\ ,$$

donde g_i es una función que sólo depende de x_i para $i = 1, \ldots, n$.

Igualmente, la independencia de n variables aleatorias se transmite a sus funciones. Si n variables son independientes, y construimos funciones suyas de forma que cada variable no intervenga más que, a lo sumo, en una de las funciones, dichas funciones serán, también, variables aleatorias independientes. Enunciamos el resultado sólo para dos funciones, aunque es válido para una cantidad cualquiera.

Proposición 5.26 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional con X_1, ..., X_n variables independientes. Sean $U = g(X_{i_1}, \ldots, X_{i_l})$ y $V = h(X_{j_1}, \ldots, X_{j_m})$ con g y h funciones reales tales que U y V sean variables aleatorias. Entonces, si $\{i_1, \ldots, i_l\}$ y $\{j_1, \ldots, j_m\}$ son disjuntos, esto es, si ninguna variable interviene en ambas funciones, U y V son independientes.*

Asimismo, para variables discretas o continuas, podríamos caracterizar la independencia por la igualdad entre las distribuciones condicionadas y las marginales. Es decir, $X_1, \ldots, X_n$ *son independientes si y sólo si, para dos grupos disjuntos cualesquiera de variables,* $(X_{i_1}, \ldots, X_{i_l})$ *y* $(X_{j_1}, \ldots, X_{j_m})$, *la distribución de* $(X_{i_1}, \ldots, X_{i_l})$ *condicionada a una realización de* $(X_{j_1}, \ldots, X_{j_m})$, *coincide con la distribución de* $(X_{i_1}, \ldots, X_{i_l})$.

5.7 Transformaciones de una variable n-dimensional

En ciertas situaciones, se conoce la distribución de probabilidad de una variable aleatoria n-dimensional, $(X_1, \ldots, X_n)$, y nos interesa estudiar la de otra u otras variables que dependen funcionalmente de aquélla. Así, por ejemplo, si $(X_1, \ldots, X_n)$ está formada por las n variables aleatorias que representan los tiempos que tardan en reaccionar n consumidores a un impulso, puede interesarnos estudiar la variable bidimensional (U, V), donde $U = \min(X_1, \ldots, X_n)$ y $V = \max(X_1, \ldots, X_n)$, es decir, el tiempo mínimo y máximo de reacción; asimismo, puede interesar la amplitud del intervalo de reacción, $Z = \max(X_1, \ldots, X_n) - \min(X_1, \ldots, X_n)$.

Analicemos detenidamente el problema para variables aleatorias bidimensionales y, como ya es habitual, generalizaremos posteriormente los resultados obtenidos a variables n-dimensionales.

Sea (Ω, α, p) un espacio probabilístico y (X, Y) una variable aleatoria bidimensional sobre dicho espacio. Sean m funciones, $h_1, \ldots, h_m$, de $\mathbb{R}^2$ en $\mathbb{R}$. Podemos considerar las m aplicaciones compuestas $U_1 = h_1(X, Y), \ldots, U_m = h_m(X, Y)$ y, por tanto, la aplicación de Ω en $\mathbb{R}^m$

$$(U_1, \ldots, U_m) = (h_1(X, Y), \ldots, h_m(X, Y)) .$$

Esquemáticamente,

$$\Omega \xrightarrow{(X,Y)} \mathbb{R}^2 \qquad \text{y} \qquad \mathbb{R}^2 \xrightarrow[(i=1,\ldots,m)]{h_i} \mathbb{R}$$

$$\Omega \xrightarrow{(X,Y)} (X,Y)(\Omega) \xrightarrow{(h_1,\ldots,h_m)} \mathbb{R}^m$$

$$(U_1, \ldots, U_m) = (h_1(X, Y), \ldots, h_m(X, Y))$$

Es decir, si ω es un punto de Ω,

$$\omega \overset{(X,Y)}{\longrightarrow} (X(\omega), Y(\omega)) \overset{(h_1,\ldots,h_m)}{\longrightarrow} (h_1(X(\omega), Y(\omega)), \ldots, h_m(X(\omega), Y(\omega)))$$

Obsérvese que, aunque $h_1, \ldots, h_m$ no estén definidas en todo el plano, lo realmente importante es que estén definidas en $(X, Y)(\Omega)$, es decir, que para todo

valor (x, y) de la variable bidimensional, existan $h_1(x, y), \ldots, h_m(x, y)$, con lo que la aplicación $(U_1, \ldots, U_m)$ estará definida para todo punto de Ω.

Para que $(U_1, \ldots, U_m)$ sea una variable aleatoria m-dimensional, es necesario y suficiente que sus componentes sean variables aleatorias. Ahora bien, una componente, $U_i = h_i(X, Y)$, será o no variable aleatoria dependiendo de la naturaleza de la función h_i, que es una aplicación de $\mathbb{R}^2$ en $\mathbb{R}$. Aunque no formalizaremos el problema, que es bastante complicado, diremos que existen funciones h_i de $\mathbb{R}^2$ en $\mathbb{R}$ tales que $h_i(X, Y)$ no es una variable aleatoria. No obstante, el problema es más bien formal y no tiene relevancia práctica. Para todas las funciones de uso habitual (exponenciales, logaritmos, funciones trigonométricas, funciones racionales o composiciones de varias de ellas) se puede demostrar que $U_i = h_i(X, Y)$ es una variable aleatoria siempre que, como antes decíamos, la aplicación compuesta U_i esté definida, esto es, que h_i esté definida al menos en $(X, Y)(\Omega)$.

El segundo problema que se plantea, en relación con las transformaciones de variables aleatorias, es el de la obtención de la distribución de probabilidades de la variable transformada. Veamos que, en general, puede obtenerse una expresión de la función de distribución de esta última en función de la probabilidad asociada a la variable original, (X, Y), expresión que, con todo, resulta de poca utilidad en los casos prácticos.

Sea $(u_1, \ldots, u_m)$ un punto de $\mathbb{R}^m$. Entonces,

$$\begin{aligned} F_{(U_1,\ldots,U_m)}(u_1, \ldots, u_m) &= p[U_1 \leq u_1, \ldots, U_m \leq u_m] = \\ &= p[h_1(X, Y) \leq u_1, \ldots, h_m(X, Y) \leq u_m] \,. \end{aligned}$$

Pero el suceso $[h_i(X, Y) \leq u_i]$ coincide con $[(X, Y) \in h_i^{-1}(-\infty, u_i]]$ para $i = 1, \ldots, m$, donde h_i^{-1} es la imagen recíproca de h_i. Por tanto,

$$\begin{aligned} F_{(U_1,\ldots,U_m)}(u_1, \ldots, u_m) &= p[(X, Y) \in h_1^{-1}(-\infty, u_1], \ldots, (X, Y) \in h_m^{-1}(-\infty, u_m]] = \\ &= p[(X, Y) \in \cap_{i=1}^m h_i^{-1}(-\infty, u_i]] = \\ &= p_{(X,Y)}[\cap_{i=1}^m h_i^{-1}(-\infty, u_i]] \end{aligned}$$

donde $p_{(X,Y)}$ es la probabilidad inducida por la variable bidimensional (X, Y).

Cuando la variable aleatoria bidimensional es discreta o continua se obtienen resultados más sencillos y útiles. Veámoslos a continuación.

5.7.1 Cambio de variable para variables bidimensionales discretas

Proposición 5.27 *Sea (X, Y) una variable aleatoria bidimensional discreta y sean g y h dos aplicaciones de $\mathbb{R}^2$ en $\mathbb{R}$ cualesquiera. Sea (U, V) la transformada de (X, Y) por (g, h), es decir, $U = g(X, Y)$ y $V = h(X, Y)$; entonces, (U, V) es una variable bidimensional discreta y su función de probabilidad, para cualquier valor, (u_k, v_l), de U y V, vale*

$$p[U = u_k, V = v_l] = \sum p[X = x_i, Y = y_j] \,,$$

donde el sumatorio se extiende a los valores (x_i, y_j) de (X, Y) tales que $g(x_i, y_j) = u_k$ y $h(x_i, y_j) = v_l$.

Demostración. Puesto que (X,Y) es discreta, los valores (x_i,y_j) de (X,Y) son, a lo sumo, una infinidad numerable.

Por otro lado,

$$[U \leq u] = [g(X,Y) \leq u] = \bigcup_{(x_i,y_j) \mid g(x_i,y_j) \leq u} [X = x_i, Y = y_j]$$

pertenecerá a α, pues se trata de la unión de una colección finita o infinita numerable de sucesos, $[X = x_i] \cap [Y = y_j]$, que pertenecen a α por ser X e Y variables aleatorias. Por tanto U es una variable aleatoria.

Como g es una aplicación, cada punto (x_i,y_j) tiene una única imagen, $g(x_i,y_j)$, y, por consiguiente, el conjunto de valores de U será, a lo sumo, infinito numerable. En consecuencia, U es discreta.

Un razonamiento análogo demostraría que V es también una variable aleatoria discreta, con lo cual (U,V) es una variable aleatoria bidimensional discreta.

Obtengamos su función de probabilidad:

$$p[U = u_k, V = v_l] = p[g(X,Y) = u_k, h(X,Y) = v_l] = \sum p[X = x_i, Y = y_j] \,,$$

donde el sumatorio se extiende a los valores (x_i,y_j) de (X,Y) tales que $g(x_i,y_j) = u_k$ y $h(x_i,y_j) = v_l$. ♣

Si el número de variables transformadas no es dos, sino una cantidad m cualquiera, el resultado anterior se demuestra igualmente. Así, *si (X,Y) es una variable aleatoria bidimensional discreta y $h_1,\ldots,h_m$ son aplicaciones de $\mathbb{R}^2$ en $\mathbb{R}$, las variables transformadas $U_1 = h_1(X,Y),\ldots,U_m = h_m(X,Y)$ son discretas y su función de probabilidad conjunta, para cualquier valor, $(u_1,\ldots,u_m)$, de $(U_1,\ldots,U_m)$, vale*

$$p[U_1 = u_1,\ldots,U_m = u_m] = \sum p[X = x_i, Y = y_j] \,,$$

donde el sumatorio se extiende a los valores (x_i,y_j) de (X,Y) tales que

$$h_1(x_i,y_j) = u_1,\ldots,h_m(x_i,y_j) = u_m \,.$$

En particular, si estamos interesados en una única transformación, $U = g(X,Y)$, la función de probabilidad de U se obtiene a partir de la de (X,Y) como

$$p[U = u_k] = \sum_{(x_i,y_j) \mid g(x_i,y_j) = u_k} p[X = x_i, Y = y_j] \,.$$

En definitiva, la obtención de la función de probabilidad de la variable aleatoria transformada, sea cual sea su dimensión, es inmediata a partir de la de (X,Y). Obsérvese, con todo, que deben conocerse previamente los valores que toma la variable transformada, que son los valores para los que calcularemos posteriormente la función de probabilidad.

Ejemplo 5.23. Sean X e Y variables aleatorias discretas cuya función de probabilidad conjunta es

$$f(x,y) = \frac{6}{37}\left(\frac{1}{x} + \frac{1}{y}\right) \quad \text{para } x = 1,2,3 \text{ e } y = 2,3.$$

Se desea hallar la función de probabilidad conjunta de $U = X \cdot Y$ y $V = X + Y$.

Escribamos la función de probabilidad conjunta de (X, Y), y sus marginales, en una tabla de doble entrada:

$X \backslash Y$	2	3	
1	9/37	8/37	17/37
2	6/37	5/37	11/37
3	5/37	4/37	9/37
	20/37	17/37	1

Multiplicando cada valor de X por cada valor de Y obtenemos los cinco valores diferentes de $U = X \cdot Y$, a saber, 2, 3, 4, 6 y 9. Asimismo, sumando cada valor de X con cada uno de los valores de Y, obtenemos los cuatro valores diferentes de $V = X + Y$, esto es, 3, 4, 5 y 6. La función de probabilidad conjunta de (U, V) se calcula para cada uno de los veinte pares, en principio, de valores posibles de la variable bidimensional. Por ejemplo,

$$p[U = 2, V = 3] = p[X = 1, Y = 2] = \frac{9}{37}$$

o

$$p[U = 6, V = 5] = p[X = 2, Y = 3] + p[X = 3, Y = 2] = \frac{5}{37} + \frac{5}{37} = \frac{10}{37}\,.$$

Todos los valores de la función de probabilidad de (U, V) se recogen en la siguiente tabla de doble entrada.

$U \backslash V$	3	4	5	6	
2	9/37	0	0	0	9/37
3	0	8/37	0	0	8/37
4	0	6/37	0	0	6/37
6	0	0	10/37	0	10/37
9	0	0	0	4/37	4/37
	9/37	14/37	10/37	4/37	1

En los márgenes de la tabla aparecen las funciones de probabilidad marginales de U (última columna) y V (última fila) que podrían haber sido obtenidas directamente. ◁

5.7.2 Cambio de variable para variables bidimensionales continuas

Si (X, Y) es una variable aleatoria continua, la técnica utilizada en el caso discreto no es aplicable a la hora de obtener la función de densidad conjunta de la variable transformada en función de la de (X, Y) (puede, incluso, que la transformada no sea continua). La función de densidad de una variable aleatoria continua bidimensional no es una probabilidad, sino el integrando de la integral que define la función de distribución conjunta o, dicho de otra forma, la segunda derivada parcial de la función de distribución con respecto a cada una de las variables.

Ejemplo 5.24. Sea (X, Y) una variable aleatoria con función de densidad

$$f_{(X,Y)}(x, y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Obtengamos la distribución de (U,V), con $U = \min(X,Y)$ y $V = \max(X,Y)$. Su función de distribución vale

$$F_{(U,V)}(u,v) = p[U \leq u, V \leq v] = p[\min(X,Y) \leq u, \max(X,Y) \leq v] \,.$$

Ahora bien,

$$[\max(X,Y) \leq v] = [\min(X,Y) \leq u, \max(X,Y) \leq v] \cup [\min(X,Y) > u, \max(X,Y) \leq v]$$

y los dos sucesos que se unen son disjuntos. Por tanto,

$$\begin{aligned} p[\max(X,Y) \leq v] \quad &= p[\min(X,Y) \leq u, \max(X,Y) \leq v] + \\ &\quad + p[\min(X,Y) > u, \max(X,Y) \leq v] \,, \end{aligned}$$

de donde resulta:

$$\begin{aligned} F_{(U,V)}(u,v) \quad &= \quad p[\min(X,Y) \leq u, \max(X,Y) \leq v] = \\ &= \quad p[\max(X,Y) \leq v] - p[\min(X,Y) > u, \max(X,Y) \leq v] \,. \end{aligned}$$

Calculemos las dos probabilidades que aparecen a la derecha de la última igualdad. Para la primera, obsérvese que

$$[\max(X,Y) \leq v] = [X \leq v, Y \leq v] \,,$$

y, por tanto,

$$p[\max(X,Y) \leq v] = p[X \leq v, Y \leq v] = p[X \leq v] \cdot p[Y \leq v] = F_X(v) \cdot F_Y(v)$$

donde la penúltima igualdad resulta de la independencia de X e Y que el lector puede comprobar con sencillez. Para calcular la segunda de las probabilidades, obsérvese que, si u es menor que v,

$$[\min(X,Y) > u, \max(X,Y) \leq v] = [u < X \leq v, u < Y \leq v]$$

y, por la independencia de X e Y,

$$\begin{aligned} p[\min(X,Y) > u, \max(X,Y) \leq v] \quad &= \quad p[u < X \leq v, u < Y \leq v] = \\ &= \quad p[u < X \leq v] \cdot p[u < Y \leq v] = \\ &= \quad [F_X(v) - F_X(u)] \cdot [F_Y(v) - F_Y(u)] \,. \end{aligned}$$

Evidentemente, si u es mayor o igual que v,

$$[\min(X,Y) > u, \max(X,Y) \leq v] = \emptyset$$

y, por tanto, su probabilidad es cero.

Resumiendo,

$$F_{(U,V)}(u,v) = \begin{cases} F_X(v) \cdot F_Y(v) & u \geq v \\ F_X(v) \cdot F_Y(v) - [F_X(v) - F_X(u)] \cdot [F_Y(v) - F_Y(u)] & u < v \end{cases}$$

Las funciones de distribución de X e Y, que se pueden obtener fácilmente, valen:

$$F_X(x) = F_Y(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Teniendo en cuenta el campo de variación de la variable transformada (véase la figura 5.19) y sustituyendo los valores de las funciones de distribución de X e Y en la expresión obtenida de $F_{(U,V)}(u,v)$, resulta:

$$F_{(U,V)}(u,v) = \begin{cases} 0 & \text{si } u < 0 \quad \text{o} \quad v < 0 \\ v^2 & \text{si } 0 \le u \le 1,\ 0 \le v \le 1 \text{ con } u \ge v \\ v^2 - (v-u)^2 & \text{si } 0 \le u \le 1,\ 0 \le v \le 1 \text{ con } u < v \\ 1 - (1-u)^2 & \text{si } 0 \le u \le 1,\ v > 1 \\ v^2 & \text{si } u > 1,\ 0 \le v \le 1 \\ 1 & \text{si } u > 1,\ v > 1 \end{cases}$$

FIG. 5.19.

La función de densidad conjunta de (U,V),

$$f_{(U,V)}(u,v) = \begin{cases} 2 & \text{si } 0 < u < v < 1 \\ 0 & \text{en el resto} \end{cases}$$

se obtiene derivando $F_{(U,V)}$ con respecto a ambas variables. ◁

En las transformaciones de variables aleatorias continuas no hay garantías ni de que las variables transformadas sean continuas, ni de que, incluso en el caso de que lo sean, su distribución conjunta sea también continua. En bastantes casos de interés, la siguiente proposición ayuda a resolver el problema.

Proposición 5.28 *Sea (X,Y) una variable aleatoria bidimensional continua cuya función de densidad conjunta es $f_{(X,Y)}(x,y)$. Sea $u = g(x,y)$, $v = h(x,y)$ una aplicación inyectiva de $\mathbb{R}^2$ en $\mathbb{R}^2$ tal que existe, en su conjunto imagen, la transformación inversa $x = g^{-1}(u,v)$, $y = h^{-1}(u,v)$. Supongamos que existen, y son continuas, las derivadas parciales $\frac{\partial x}{\partial u}$, $\frac{\partial x}{\partial v}$, $\frac{\partial y}{\partial u}$ y $\frac{\partial y}{\partial v}$, y que el determinante jacobiano*

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

es distinto de cero. Sea (U,V) la transformada de (X,Y) por (g,h), es decir, $U = g(X,Y)$ y $V = h(X,Y)$; entonces, (U,V) es una variable aleatoria bidimensional continua con función de densidad

$$f_{(U,V)}(u,v) = \begin{cases} f_{(X,Y)}(g^{-1}(u,v), h^{-1}(u,v)) \cdot |J| & si\ (u,v) \in D \\ 0 & si\ (u,v) \notin D \end{cases}$$

donde $D = (g,h)(I\!R^2)$, *esto es,*

$$D = \{(g(x,y), h(x,y)), \textit{para todo } (x,y) \in I\!R^2\} .$$

Demostración. Dado un punto (a,b) del plano, calculemos la función de distribución de (U,V) en dicho punto,

$$F_{(U,V)}(a,b) = p[U \leq a, V \leq b] .$$

El suceso $[U \leq a, V \leq b]$ puede escribirse como

$$[U \leq a, V \leq b] = [g(X,Y) \leq a, h(X,Y) \leq b] = [(X,Y) \in S]$$

donde S es el subconjunto del plano

$$S = \{(x,y) \text{ tales que } g(x,y) \leq a, h(x,y) \leq b\} .$$

En definitiva,

$$F_{(U,V)}(a,b) = p[(X,Y) \in S] = \int\int_S f_{(X,Y)}(x,y)dxdy .$$

Si en esta última integral hacemos el cambio de variable $x = g^{-1}(u,v)$, $y = h^{-1}(u,v)$, como (g^{-1}, h^{-1}) es la transformación inversa de (g,h), el recinto S se transformará en

$$T = \{(u,v) \text{ tales que } u \leq a, v \leq b\}$$

y, aplicando a la integral anterior el teorema del cambio de variable para integrales dobles,

$$\int\int_S f_{(X,Y)}(x,y)dxdy = \int\int_T f_{(X,Y)}(g^{-1}(u,v), h^{-1}(u,v)) \cdot |J| \; dudv ,$$

donde $|J|$ es el valor absoluto del determinante jacobiano que describimos en el enunciado. Por tanto,

$$F_{(U,V)}(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{(X,Y)}(g^{-1}(u,v), h^{-1}(u,v)) \cdot |J| \; dudv ,$$

para todo punto (a,b) del plano, con lo que la función integrando es la función de densidad conjunta de (U,V). Si la aplicación (g,h) no fuera suprayectiva, es decir, si $D = (g,h)(\mathbb{R}^2)$ es un subconjunto del plano que no coincide con el plano, entonces la función de densidad conjunta de (U,V) valdrá cero para los puntos (u,v) que no pertenezcan a D ya que, obviamente, (U,V) no tomará valores fuera de D.

En definitiva,

$$f_{(U,V)}(u,v) = \begin{cases} f_{(X,Y)}(g^{-1}(u,v), h^{-1}(u,v)) \cdot |J| & \text{si } (u,v) \in D \\ 0 & \text{si } (u,v) \notin D \end{cases}$$

que, como puede comprobarse, es realmente la función de densidad de una variable aleatoria bidimensional continua. ♣

Ejemplo 5.25. Sea (X,Y) una variable aleatoria con función de densidad conjunta

$$f_{(X,Y)}(x,y) = \begin{cases} e^{-(x+y)} & \text{si } x > 0,\ y > 0 \\ 0 & \text{en el resto} \end{cases}$$

Vamos a obtener la función de densidad conjunta de (U, V), donde $U = X + Y$ y $V = X - Y$.

Se cumplen las condiciones que permiten aplicar el teorema del cambio de variable. La transformación del plano en sí mismo es

$$u = x + y \qquad v = x - y$$

cuya inversa

$$x = \frac{u+v}{2} \qquad y = \frac{u-v}{2}$$

se obtiene sin dificultad. Las derivadas parciales de la transformación inversa existen y son continuas (de hecho son constantes), y el determinante jacobiano resulta

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

que es distinto de cero.

En definitiva, la función de densidad conjunta de (U, V) es

$$f_{(U,V)}(u,v) = f_{(X,Y)}(x(u,v), y(u,v)) \cdot |J| = \begin{cases} \frac{1}{2} \cdot e^{-u} & \text{si } \frac{u+v}{2} > 0,\ \frac{u-v}{2} > 0 \\ 0 & \text{en el resto} \end{cases}$$

FIG. 5.20.

Obsérvese que, como pone en evidencia la figura 5.20,

$$u > 0 \qquad -u < v < u$$

es una forma más sencilla de expresar el recinto en el que $f_{(U,V)}$ es distinta de cero, esto es, el conjunto de valores de la variable bidimensional (U, V). ◁

Como puede verse en la demostración de la proposición anterior, las condiciones de inyectividad de la aplicación, existencia y continuidad de las derivadas parciales y no anulación del determinante jacobiano, se imponen para que pueda aplicarse el teorema del cambio de variable para integrales dobles.

Las derivadas parciales que componen el determinante jacobiano son las de la transformación inversa; por ejemplo, $\frac{\partial x}{\partial u}$ es la derivada parcial $\frac{\partial g^{-1}(u,v)}{\partial u}$. Recuérdese que, si ello resultara más sencillo, puede obtenerse el determinante jacobiano de la transformación directa,

$$J^* = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix},$$

ya que ambos determinantes están ligados por la relación

$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{1}{\frac{\partial(u,v)}{\partial(x,y)}}.$$

Por otro lado, aunque el teorema exige que la función (g,h) esté definida en todo el plano, lo único realmente importante es que esté definida en el campo de variación de (X,Y), esto es, en el conjunto

$$\{(x,y) \text{ tales que } f_{(X,Y)}(x,y) \neq 0\}\ .$$

Veámoslo en el siguiente ejemplo.

Ejemplo 5.26. Sea (X,Y) una variable aleatoria cuya función de densidad es

$$f_{(X,Y)}(x,y) = \begin{cases} x+y & \text{si } 0<x<1,\ 0<y<1 \\ 0 & \text{en el resto} \end{cases}$$

Obtengamos la función de densidad conjunta de $U = \ln(X)$ y $V = X+Y$.

Obsérvese que la transformación $u = \ln(x)$ no está definida en los puntos (x,y) tales que x sea negativo. Pero como la variable aleatoria (X,Y) sólo toma valores positivos para X (en concreto, los del intervalo $(0,1)$), el teorema del cambio de variable es aplicable. Escribamos la transformación

$$u = \ln(x) \qquad v = x+y\ .$$

La transformación inversa es

$$x = e^u \qquad y = v - e^u$$

y sus derivadas parciales son

$$\frac{\partial x}{\partial u} = e^u, \quad \frac{\partial x}{\partial v} = 0, \quad \frac{\partial y}{\partial u} = -e^u \quad \text{y} \quad \frac{\partial y}{\partial v} = 1\ .$$

El determinante jacobiano, $J = e^u$, es siempre positivo, por lo cual coincide con su valor absoluto. La función de densidad conjunta de (U,V) será

$$f_{(U,V)}(u,v) = \begin{cases} (e^u + v - e^u)\cdot e^u & \text{si } 0<e^u<1,\ 0<v-e^u<1 \\ 0 & \text{en el resto} \end{cases}$$

El conjunto de valores de (U,V), esto es, el subconjunto del plano en el que su función de densidad conjunta es distinta de cero, puede escribirse de forma más sencilla. Así,

$$f_{(U,V)}(u,v) = \begin{cases} v\cdot e^u & \text{si } -\infty<u<0,\ e^u<v<1+e^u \\ 0 & \text{en el resto} \end{cases}$$

expresión obtenida sin más que simplificar la función de densidad. ◁

El teorema del cambio de variable permite hallar la función de densidad de una variable bidimensional (U,V), transformada de otra variable bidimensional (X,Y). Como hemos visto, (U,V) es continua si la transformación (g,h) cumple las condiciones de la proposición anterior. Pero el teorema del cambio de variable del

cálculo integral transforma integrales dobles en integrales dobles y no sirve si, por ejemplo, queremos estudiar la distribución de una variable aleatoria tridimensional (U, V, W), donde $U = g(X, Y)$, $V = h(X, Y)$ y $W = l(X, Y)$, a diferencia de lo que ocurría cuando las variables eran discretas.

El hecho es que si (X, Y) es continua, (U, V, W) no puede serlo, y, por tanto, no tiene sentido hablar de su función de densidad. Como antes dijimos, lo único que se puede hacer en estos casos es obtener la función de distribución de (U, V, W), y utilizarla para obtener probabilidades referentes a la variable transformada.

Cuando sólo nos interesa una variable aleatoria que se obtiene por transformación de la variable aleatoria bidimensional, esto es, si estamos interesados en estudiar $U = g(X, Y)$, el teorema del cambio de variable sigue sin ser aplicable. No obstante, en este caso puede recurrirse a la siguiente técnica: construimos otra transformación de apoyo, $V = h(X, Y)$, de forma que (g, h) verifique las condiciones de la proposición anterior (usualmente se elige una función, h, sencilla, como por ejemplo $V = X$ o $V = Y$). Aplicando el teorema del cambio de variable, se halla la función de densidad conjunta de (U, V); integrando esta última en el campo de variación de V, se obtiene la función de densidad de U.

Ejemplo 5.27. Sea, de nuevo, la variable aleatoria (X, Y) cuya función de densidad es

$$f_{(X,Y)}(x, y) = \begin{cases} x + y & 0 < x < 1, 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Vamos a hallar la distribución de la variable $U = X \cdot Y$. Podemos proceder de dos maneras diferentes: bien completando la transformación para que se pueda aplicar el teorema del cambio de variable, bien calculando directamente la función de distribución de la variable transformada.

Completemos la transformación introduciendo una nueva variable, $V = X$. La transformación

$$u = x \cdot y \qquad v = x$$

y su inversa

$$x = v \qquad y = \frac{u}{v}$$

verifican las condiciones del teorema del cambio de variable, siendo el jacobiano:

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} 0 & 1 \\ 1/v & -u/v^2 \end{vmatrix} = -\frac{1}{v}\,.$$

Entonces,

$$f_{(U,V)}(u, v) = f_{(X,Y)}(x(u, v), y(u, v)) \cdot |J| = \begin{cases} (v + \frac{u}{v}) \cdot \frac{1}{v} & 0 < v < 1, 0 < \frac{u}{v} < 1 \\ 0 & \text{en el resto} \end{cases}$$

Es decir,

$$f_{(U,V)}(u, v) = \begin{cases} 1 + \frac{u}{v^2} & 0 < u < v < 1 \\ 0 & \text{en el resto} \end{cases}$$

Por tanto, la función de densidad marginal de U será:

$$f_U(u) = \int_{v=u}^{1} \left(1 + \frac{u}{v^2}\right) dv = \begin{cases} 2(1 - u) & 0 < u < 1 \\ 0 & \text{en el resto} \end{cases}$$

Otra posibilidad es plantear directamente la función de distribución de U, es decir,

$$F_U(u) = p[U \leq u] = p[X \cdot Y \leq u] = \int\int_S f_{(X,Y)}(x,y)dxdy \ ,$$

donde S es el recinto en el que $X \cdot Y \leq u$ (véase la figura 5.21).

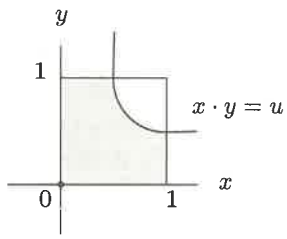


FIG. 5.21.

Por tanto,

$$F_U(u) = \begin{cases} 0 & u \leq 0 \\ \int\int_S (x+y)dxdy = 2u - u^2 & 0 < u < 1 \\ 1 & u \geq 1 \end{cases}$$

Derivando obtendríamos la función de densidad de U. ◁

Cuando la transformación no es inyectiva, o no existen las derivadas parciales, debe recurrirse al cálculo de la función de distribución conjunta de la variable transformada. A veces ésta es continua y puede deducirse, mediante derivación, su función de densidad conjunta. Veamos el siguiente ejemplo.

Ejemplo 5.28. Sea (X,Y) una variable bidimensional continua con función de densidad

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{9}{4}x^2 \cdot y^2 & \text{si } -1 < x < 1, \ -1 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Obtengamos la distribución de probabilidad de (U,V), siendo $U = X^2$ y $V = Y^2$.

La transformación del plano en sí mismo es

$$u = x^2 \qquad v = y^2 \ .$$

Esta transformación no es suprayectiva (aplica el plano en el primer cuadrante) y, además, tampoco es inyectiva (el punto $(0.25, 0.25)$, por ejemplo, es imagen de los puntos $(0.5, 0.5)$, $(-0.5, 0.5)$, $(0.5, -0.5)$ y $(-0.5, -0.5)$). Por tanto, el teorema del cambio de variable no es aplicable.

Obtengamos la función de distribución conjunta. Como X e Y varían entre -1 y 1, las variables transformadas tomarán todos los valores entre 0 y 1. En consecuencia, la

función de distribución conjunta de (U,V) se anula si u o v son negativos, y es igual a uno si u y v son ambos mayores que la unidad. Para el resto de los puntos (u,v) del plano,

$$\begin{aligned} F_{(U,V)}(u,v) &= p[U \leq u, V \leq v] = p[X^2 \leq u, Y^2 \leq v] = \\ &= p[-\sqrt{u} \leq X \leq \sqrt{u}, -\sqrt{v} \leq Y \leq \sqrt{v}] = \\ &= p[-\sqrt{u} \leq X \leq \sqrt{u}] \cdot p[-\sqrt{v} \leq Y \leq \sqrt{v}] \end{aligned}$$

donde la última igualdad se verifica porque X e Y son independientes, como comprobamos en el ejemplo 5.19.

Por tanto, si u y v están en el intervalo $(0,1)$,

$$F_{(U,V)}(u,v) = \int_{-\sqrt{u}}^{\sqrt{u}} \frac{3}{2}x^2 dx \cdot \int_{-\sqrt{v}}^{\sqrt{v}} \frac{3}{2}y^2 dy = u^{3/2} \cdot v^{3/2} .$$

Si $0 \leq u \leq 1$ y $v > 1$,

$$F_{(U,V)}(u,v) = \int_{-\sqrt{u}}^{\sqrt{u}} \frac{3}{2}x^2 dx \cdot \int_{-1}^{1} \frac{3}{2}y^2 dy = u^{3/2} ,$$

y si $u > 1$ y $0 \leq v \leq 1$,

$$F_{(U,V)}(u,v) = \int_{-1}^{1} \frac{3}{2}x^2 dx \cdot \int_{-\sqrt{v}}^{\sqrt{v}} \frac{3}{2}y^2 dy = v^{3/2} .$$

Agrupando los resultados obtenidos, la función de distribución conjunta de U y V resulta ser:

$$F_{(U,V)}(u,v) = \begin{cases} 0 & u < 0 \quad \text{o} \quad v < 0 \\ u^{3/2} \cdot v^{3/2} & 0 \leq u \leq 1, 0 \leq v \leq 1 \\ u^{3/2} & 0 \leq u \leq 1, v > 1 \\ v^{3/2} & u > 1, 0 \leq v \leq 1 \\ 1 & u > 1, v > 1 \end{cases}$$

Derivando, resulta la función de densidad:

$$f_{(U,V)}(u,v) = \begin{cases} \frac{9}{4}\sqrt{u \cdot v} & 0 \leq u \leq 1, 0 \leq v \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Obsérvese que la resolución del ejercicio se ha complicado innecesariamente. Como X e Y son independientes, U y V lo son y, por tanto, se puede obtener la función de densidad conjunta de (U,V) como producto de las marginales. Estas funciones de densidad marginales de U y V se obtendrían a partir de las de X e Y, respectivamente, utilizando el procedimiento del ejemplo 4.17. El motivo por el que hemos resuelto el ejemplo de una forma tan complicada ha sido el que el lector se familiarizase con la técnica de cambio de variable cuando la proposición 5.28 no es aplicable. ◁

Los resultados de los dos apartados anteriores se generalizan a variables aleatorias de dimensión superior.

Sea (Ω, α, p) un espacio probabilístico y $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional sobre dicho espacio. Sean m funciones, $h_1, \ldots, h_m$, de $\mathbb{R}^n$ en $\mathbb{R}$.

Entonces, podemos considerar las m aplicaciones compuestas $U_1 = h_1(X_1, \ldots, X_n)$, $\ldots$, $U_m = h_m(X_1, \ldots, X_n)$ y, por tanto, la aplicación de Ω en $\mathbb{R}^m$

$$(U_1, \ldots, U_m) = (h_1(X_1, \ldots, X_n), \ldots, h_m(X_1, \ldots, X_n)) .$$

Aunque $h_1, \ldots, h_m$ no estén definidas en todo $\mathbb{R}^n$, lo realmente importante es que estén definidas en $(X_1, \ldots, X_n)(\Omega)$, es decir, que para todo valor $(x_1, \ldots, x_n)$ de la variable n-dimensional, existan $h_1(x_1, \ldots, x_n), \ldots, h_m(x_1, \ldots, x_n)$, con lo que la aplicación compuesta $(U_1, \ldots, U_m)$ estará definida para todo punto de Ω.

Como vimos antes para variables bidimensionales, puede obtenerse una expresión de la función de distribución de la variable transformada en función de la probabilidad inducida por la variable original. En efecto, sea $(u_1, \ldots, u_m)$ un punto de $\mathbb{R}^m$. Entonces,

$$\begin{aligned} F_{(U_1,\ldots,U_m)}(u_1, \ldots, u_m) &= p[U_1 \leq u_1, \ldots, U_m \leq u_m] = \\ &= p[h_1(X_1, \ldots, X_n) \leq u_1, \ldots, h_m(X_1, \ldots, X_n) \leq u_m] = \\ &= p_{(X_1,\ldots,X_n)}[\cap_{i=1}^m h_i^{-1}(-\infty, u_i]] , \end{aligned}$$

donde $p_{(X_1,\ldots,X_n)}$ es la probabilidad inducida por la variable aleatoria n-dimensional $(X_1, \ldots, X_n)$.

Se obtienen, también en este caso, resultados más sencillos y útiles cuando la variable aleatoria n-dimensional es discreta o continua. Los resultados se demuestran como en el caso bidimensional.

Proposición 5.29 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional discreta y $h_1, \ldots, h_m$, aplicaciones de $\mathbb{R}^n$ en $\mathbb{R}$ cualesquiera. Sean $U_1 = h_1(X_1, \ldots, X_n)$, $\ldots$, $U_m = h_m(X_1, \ldots, X_n)$; entonces, $(U_1, \ldots, U_m)$ es una variable m-dimensional discreta y su función de probabilidad, para un valor cualquiera $(u_1, \ldots, u_m)$ de $(U_1, \ldots, U_m)$, vale*

$$p[U_1 = u_1, \ldots, U_m = u_m] = \sum p[X_1 = x_1, \ldots, X_n = x_n] ,$$

donde el sumatorio se extiende a los valores $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$ tales que

$$h_1(x_1, \ldots, x_n) = u_1, \ldots, h_m(x_1, \ldots, x_n) = u_m .$$

Se particulariza, asimismo, el resultado a una única transformación,

$$U = g(X_1, \ldots, X_n) .$$

La función de probabilidad de U se obtiene a partir de la de $(X_1, \ldots, X_n)$ como

$$p[U = u_k] = \sum_{(x_1,\ldots,x_n) \mid g(x_1,\ldots,x_n)=u_k} p[X_1 = x_1, \ldots, X_n = x_n] .$$

Si la variable $(X_1, \ldots, X_n)$ es continua, en muchos casos el resultado siguiente permite hallar la distribución de probabilidades de una variable n-dimensional que se obtenga por transformación de aquélla.

Proposición 5.30 *Sea $(X_1,\ldots,X_n)$ una variable aleatoria n-dimensional continua cuya función de densidad conjunta es $f_{(X_1,\ldots,X_n)}$. Sea*

$$u_1 = h_1(x_1,\ldots,x_n),\ldots,u_n = h_n(x_1,\ldots,x_n)\ ,$$

una aplicación inyectiva de $\mathbb{R}^n$ en $\mathbb{R}^n$ tal que existe, en su conjunto imagen, la transformación inversa

$$x_1 = h_1^{-1}(u_1,\ldots,u_n),\ldots,x_n = h_n^{-1}(u_1,\ldots,u_n)\ .$$

Supongamos que existen y son continuas las derivadas parciales $\frac{\partial x_1}{\partial u_1}$, ..., $\frac{\partial x_1}{\partial u_n}$, ..., $\frac{\partial x_n}{\partial u_1}$, ..., $\frac{\partial x_n}{\partial u_n}$, y que el determinante jacobiano

$$J = \frac{\partial(x_1,\ldots,x_n)}{\partial(u_1,\ldots,u_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial u_1} & \cdots & \frac{\partial x_n}{\partial u_n} \end{vmatrix}$$

es distinto de cero.

Sea $(U_1,\ldots,U_n)$ la transformada de $(X_1,\ldots,X_n)$ por $(h_1,\ldots,h_n)$, es decir,

$$U_1 = h_1(X_1,\ldots,X_n),\ldots,U_n = h_n(X_1,\ldots,X_n)\ ;$$

entonces, $(U_1,\ldots,U_n)$ es una variable aleatoria n-dimensional continua con función de densidad

$$f_{(U_1,\ldots,U_n)}(u_1,\ldots,u_n) = f_{(X_1,\ldots,X_n)}(h_1^{-1}(u_1,\ldots,u_n),\ldots,h_n^{-1}(u_1,\ldots,u_n))\cdot|J|$$

si $(u_1,\ldots,u_n)\in D$, anulándose en el resto, donde $D = (h_1,\ldots,h_n)(\mathbb{R}^n)$, esto es,

$$D = \{(h_1(x_1,\ldots,x_n),\ldots,h_n(x_1,\ldots,x_n)),\quad \textit{para todo}\quad (x_1,\ldots,x_n)\in\mathbb{R}^n\}\ .$$

Obsérvese que, como dijimos anteriormente en el caso bidimensional, aunque el teorema exige que la función $(h_1,\ldots,h_n)$ esté definida en todo el plano, lo único realmente importante es que esté definida en el campo de variación de $(X_1,\ldots,X_n)$, esto es, en el conjunto

$$\{(x_1,\ldots,x_n)\text{ tales que }f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)\neq 0\}\ .$$

El resultado anterior permite únicamente obtener la función de densidad de una variable n-dimensional transformada de otra de la misma dimensión. Si la dimensión de la variable transformada es superior a n, ésta no puede ser continua. Si es inferior a n, debemos completar la transformación tal y como comentamos en el caso bidimensional y luego obtener la distribución marginal correspondiente.

Cuando la transformación no es inyectiva, o no existen las derivadas parciales, debe recurrirse al cálculo de la función de distribución conjunta de la variable transformada. A veces ésta es continua y puede obtenerse mediante derivación su función de densidad conjunta, aunque las dificultades de manejo de funciones de distribución de variables n-dimensionales hacen que esta posibilidad sea más formal que real.

5.8 Ejercicios propuestos

5.1. Dada una variable aleatoria (X, Y), demuéstrese que, si $B = A \times C$ donde A y C pertenecen a la σ-álgebra de Borel, β, se verifica que

$$(X, Y)^{-1}(B) \in \alpha \, .$$

5.2. Pruébese que la probabilidad inducida dada en la definición 5.3 es realmente una probabilidad.

5.3. Demuéstrese que una función G que verifique las tres condiciones planteadas en la proposición 5.2 es función de distribución de una variable aleatoria.

5.4. Tenemos una urna con tres bolas blancas y dos negras. Se realizan tres extracciones sin reemplazamiento. Sea X la variable que designa el número de bolas blancas obtenidas e Y la variable que expresa el número de bolas negras antes de obtener la primera bola blanca.

a) Hállese la distribución conjunta de X e Y.

b) Calcúlense las funciones de probabilidad marginales.

5.5. Se lanza un dado dos veces. Obténgase la distribución conjunta de las variables que designan el resultado del primer lanzamiento y el máximo resultado de los dos lanzamientos.

5.6. La junta directiva de una organización femenina está formada por 4 mujeres solteras, 3 divorciadas, 2 viudas y una casada.
Con objeto de elaborar el texto para unos folletos de carácter divulgativo, se elige al azar una comisión formada por 3 personas.

a) Hállese la función de probabilidad conjunta de X, número de mujeres solteras, e Y, número de mujeres viudas, incluidas en dicha comisión.

b) Calcúlese la probabilidad de que haya al menos una mujer viuda en la comisión.

c) ¿Cuál es la probabilidad de que el número de solteras que integran la comisión sea superior al de viudas?

5.7. Demuéstrese que una función g que verifique las dos condiciones planteadas en el apartado *2* de la proposición 5.5 es función de probabilidad de una variable aleatoria discreta.

5.8. Razónese si las siguientes funciones son, o no, funciones de densidad:

a) $f(x, y) = e^{-x}$ si $0 < y < x < \infty$ y cero en el resto.

b) $f(x, y) = 2(1 + x \cdot y)$ si $|x| \leq 1$, $|y| \leq 1$ y cero en el resto.

c) $f(x, y) = \frac{1}{5} x \cdot y^2$ si $0 < x < y < 1$ y cero en el resto.

5.9. Demuéstrese que la función

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \quad \text{o} \quad y < 0 \\ x \cdot y & \text{si } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ x & \text{si } 0 \leq x \leq 1,\ y > 1 \\ y & \text{si } x > 1,\ 0 \leq y \leq 1 \\ 1 & \text{si } x > 1,\ y > 1 \end{cases}$$

es función de distribución de una variable aleatoria bidimensional continua. Hállense las funciones de distribución marginales de X y de Y.

5.10. Dada una variable aleatoria bidimensional continua (X,Y), cuya función de densidad es $f(x,y)$, pruébese que,

$$\lim_{\substack{h\to 0 \\ k\to 0}} \frac{p\left[x - \frac{h}{2} \leq X \leq x + \frac{h}{2}, y - \frac{k}{2} \leq Y \leq y + \frac{k}{2}\right]}{h \cdot k} = f(x,y)\ .$$

(Sugerencia: Téngase en cuenta la observación **3** de la sección 4.5 y aplíquese el teorema de los incrementos finitos a las funciones $\phi(x) = F(x, y+\frac{k}{2}) - F(x, y-\frac{k}{2})$ y $\psi(y) = \frac{\partial}{\partial x}F(x,y)$, sucesivamente.)

5.11. Demuéstrese que una función g que verifique las dos condiciones planteadas en la proposición 5.6 es función de densidad de una variable aleatoria continua.

5.12. Compruébese que, dada una variable aleatoria bidimensional continua con función de densidad $f(x,y)$, la función $g(x) = \int_{-\infty}^{\infty} f(x,y)dy$ es realmente una función de densidad.

5.13. El número de errores, X, que una mecanógrafa comete por página, es una variable aleatoria con función de probabilidad:

$$p[X = x] = e^{-2} \cdot \frac{2^x}{x!} \quad x = 0, 1, \ldots$$

Si en una página tiene x erratas, el número de minutos, Y, que emplea en revisar y corregir dicha página, es una variable con distribución:

$$p[Y = y \mid X = x] = \begin{cases} 1/5 & \text{si} \quad y = 1 + x \\ 3/5 & \text{si} \quad y = 2 + x \\ 1/5 & \text{si} \quad y = 3 + x \end{cases}$$

a) Encuéntrese la probabilidad de que se necesiten 4 minutos para revisar y corregir una página elegida al azar.

b) Si se han empleado 5 minutos en la revisión y corrección de una página, ¿cuál es la probabilidad de que tuviera 3 erratas?

5.14. Se lanza un dado equilibrado. Sean dos variables aleatorias X e Y definidas por:

$$X = \begin{cases} -1 & \text{si el resultado es impar} \\ 1 & \text{si el resultado es par} \end{cases}$$

$$Y = \begin{cases} -1 & \text{si el resultado es 1, 2 o 3} \\ 0 & \text{si el resultado es 4} \\ 1 & \text{si es 5 o 6} \end{cases}$$

a) Hállense las funciones de probabilidad y de distribución conjunta de X y de Y.

b) Calcúlese $p[X+Y=0 \mid Y \leq 0]$ y $p[X=1 \mid X+Y=2]$.

5.15. Sea

$$f(x,y)=\begin{cases} \frac{1}{4}\cdot\frac{10-x}{x}\cdot\left(\frac{x}{10}\right)^{10} & x=5,6,7,8;\ y=1,2,\ldots \\ 0 & \text{en el resto} \end{cases}$$

la función de densidad conjunta de (X,Y), donde X es el número de facturas elegidas al azar e Y el tiempo, en minutos, que se tarda en revisarlas (véase el ejemplo 5.12).

a) Hállese la distribución marginal de Y.

b) Obténgase la distribución de X condicionada a que $Y=y$.

c) Calcúlese la probabilidad de que se eligieran 6 facturas, sabiendo que se emplearon más de 3 minutos en revisarlas.

5.16. Sea (X,Y) una variable aleatoria bidimensional. Si la función de densidad de X es

$$f_X(x)=\begin{cases} 4x^3 & \text{si}\quad 0<x<1 \\ 0 & \text{en el resto} \end{cases}$$

y la función de densidad de Y dado $X=x$,

$$f_{(Y\mid X=x)}(y)=\begin{cases} \frac{2y}{x^2} & 0<y<x \\ 0 & \text{en el resto} \end{cases}$$

a) Obténgase la función de densidad de (X,Y).

b) Calcúlese la distribución de la variable Y.

c) Hállese la función de desidad de X condicionada a $Y=y$.

5.17. La variable aleatoria (R,C) representa la renta y el consumo, en millones de pesetas, de un grupo de trabajadores de una cierta empresa, siendo su función de densidad:

$$f(r,c)=\begin{cases} \frac{1}{6}r^{1/2}\cdot c^{1/2} & 0<c<r<3 \\ 0 & \text{en el resto} \end{cases}$$

Calcúlense:

a) Las funciones de densidad marginales.

b) La probabilidad de que el consumo de un individuo sea superior a un millón si se sabe que su renta no es superior a 2 millones y medio.

5.18. Un novelista se ha comprometido con su editor a entregarle cada 30 días un capítulo del libro que está escribiendo; éste ha sido un método de trabajo que, desde hace muchos años, le ha dado óptimos resultados.
El número de días que emplea para redactar cada capítulo es una variable aleatoria, X, que se modeliza según la siguiente función de densidad:

$$f_X(x)=\begin{cases} \frac{1}{30} & 0\leq x\leq 30 \\ 0 & \text{en el resto} \end{cases}$$

Si tarda x días en redactar el capítulo, el tiempo total, Y, que tarda en elaborarlo (incluyendo la revisión y corrección de erratas) responde a la siguiente función de densidad:

$$f_{(Y \mid X=x)}(y) = \begin{cases} \frac{1}{30-x} & 0 \leq x \leq y \leq 30 \\ 0 & \text{en el resto} \end{cases}$$

a) Determínese la probabilidad de que emplee más de 25 días en la ejecución total del capítulo.

b) Hállese la probabilidad de que el número de días de redacción sea superior al número de días empleados en la corrección.

c) Obténgase la probabilidad de que el número de días empleado en la corrección sea inferior a 2.

d) Calcúlese la probabilidad de que invierta entre 15 y 20 días en la redacción de un capítulo si el número total de días hasta finalizarlo ha sido 25.

5.19. El número de hamburguesas, en miles de unidades, que se venden semanalmente en un *burguer* es una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} \frac{x^2}{9} & 0 < x < 3 \\ 0 & \text{en el resto} \end{cases}$$

Si en una semana se venden x hamburguesas, la cantidad de ellas que llevan queso, Y, se distribuye según la siguiente función de densidad:

$$f_{(Y \mid X=x)}(y) = \begin{cases} \frac{1}{x} & 0 < y < x < 3 \\ 0 & \text{en el resto} \end{cases}$$

a) Sabiendo que en una semana se han vendido mil hamburguesas, ¿cuál es la probabilidad de que al menos la mitad tuvieran queso?

b) Si el precio por hamburguesa con queso es de 350 pesetas, hállese la probabilidad de que los ingresos para este tipo de consumición sean superiores a 200 mil pesetas.

c) Obténgase la probabilidad de que en una semana el número vendido de hamburguesas sin queso sea superior a mil.

5.20. De cara a preparar la próxima temporada veraniega, el propietario de Pochó y Vengo-Vengo, locales habituales en las noches de una cierta ciudad costera, muestra un especial interés por el consumo de bebidas no alcohólicas.
Sean X e Y las proporciones semanales de bebidas sin alcohol sobre el total de consumiciones en Pochó y Vengo-Vengo, respectivamente. La función de densidad conjunta de estas variables es:

$$f(x,y) = \begin{cases} kx \cdot (x+y) & 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Determínese el valor de k.

b) Obténgase el porcentaje de semanas en que la proporción de bebidas alcohólicas en Pochó supera el 30 %.

c) ¿Cuál es la probabilidad de que el porcentaje de bebidas alcohólicas sea mayor en Pochó que en Vengo-Vengo?

d) Si en una semana el porcentaje de bebidas no alcohólicas supera el 50 % en Pochó, ¿cuál es la probabilidad de que ocurra lo mismo en Vengo-Vengo?

5.21. La cantidad de un cierto producto, en miles de unidades, X, que una empresa vende al mes, y su precio, en miles de pesetas por unidad, Y, tienen la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} \frac{x \cdot y}{96} & 0 < x < 4, \quad 1 < y < 5 \\ 0 & \text{en el resto} \end{cases}$$

a) Si el precio por unidad en un cierto mes es de 4 mil pesetas, hállese la probabilidad de que la cantidad vendida sea superior a 2 mil unidades.

b) Determínese el porcentaje de meses en los que el número de unidades vendidas es superior a mil y el precio por unidad inferior a 2 mil pesetas.

c) Calcúlese la probabilidad de que en un mes el ingreso total debido a la venta de este producto, sea superior a 4 millones de pesetas.

5.22. Una farmacia posee una cantidad, X, de cientos de unidades de un cierto fármaco al comienzo de cada mes. Durante el mes se venden Y cientos de unidades del referido fármaco. Si

$$f(x,y) = \begin{cases} 2/9 & 0 < x < y < 3 \\ 0 & \text{en el resto} \end{cases}$$

es la función de densidad conjunta de X e Y:

a) Demuéstrese que f es función de densidad.

b) Obténgase la función de distribución de (X,Y).

c) Hállese la probabilidad de que a final de mes se hayan vendido al menos la mitad de las unidades que había inicialmente.

d) Si se han vendido cien unidades, ¿cuál es la probabilidad de que hubiera al menos doscientas a principios de mes?

5.23. Un estudio realizado sobre las ventas en una frutería arroja los siguientes resultados:

- El 10 % de los clientes compra un kilo de manzanas y uno de naranjas.
- El 30 % de los clientes sólo un kilo de manzanas.
- El 15 % de los clientes sólo un kilo de naranjas.
- El 40 % compra manzanas y naranjas en cantidades inferiores al kilo en ambos casos.
- El 5 % restante no compra ni naranjas ni manzanas.

a) ¿Cuál es la probabilidad de que un cliente compre al menos un kilo de fruta?

b) ¿Qué probabilidad hay de que un cliente compre menos de medio kilo de manzanas y más de un cuarto de naranjas?

c) Si un cliente ha comprado, entre naranjas y manzanas, una cantidad inferior al kilo, ¿cuál es la probabilidad de que haya adquirido menos de medio kilo de manzanas?

5.24. Compruébese que la función de densidad condicionada

$$f_{(X \mid Y=y)}(x) = \frac{f(x,y)}{f_Y(y)} ,$$

es realmente una función de densidad.

5.25. Demuéstrese que, dadas dos variables aleatorias cualesquiera independientes, sus cuadrados también son independientes.

5.26. Sean X e Y variables aleatorias discretas independientes tales que X toma los valores 2, 4 y 6 e Y los valores 3 y 5. Sabiendo que $p[X = 4] = 0.1$, $p[Y = 5] = 0.6$ y $p[X = 2, Y = 5] = 0.5$, obténganse las distribuciones de probabilidad marginales, así como la distribución conjunta de dichas variables.

5.27. Sean X_i, $i = 1, 2$, variables aleatorias independientes e igualmente distribuidas tales que $p(X_i = 1) = p(X_i = -1) = 1/2$. Se considera la variable aleatoria $X_3 = X_1 \cdot X_2$. ¿Son las variables X_1, X_2 y X_3 independientes? ¿Son independientes dos a dos?

5.28. Sean X e Y variables aleatorias independientes con funciones de densidad

$$f_X(x) = \begin{cases} x \cdot e^{-x} & \text{si} \quad x > 0 \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \begin{cases} 1/2 & \text{si} \quad 0 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

Obténgase una función de densidad del par (X, Y) cuyas funciones de densidad marginales sean las anteriores.

5.29. La empresa OLIS, encargada de la fabricación y el mantenimiento de ascensores, desea mejorar la calidad de sus productos y el servicio al cliente. A tal fin, se lleva a cabo un análisis de la situación actual del que se concluye que la distribución conjunta de la variable aleatoria X, tiempo transcurrido, en días, desde la instalación de un ascensor hasta la primera avería, e Y, tiempo, en días, que tarda la empresa en acudir al lugar del problema para solucionarlo, viene dada por:

$$f(x,y) = \begin{cases} e^{-\frac{x+900y}{300}} & x > 0 \quad y > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Calcúlese la función de densidad del tiempo de funcionamiento hasta la primera avería.

b) El jefe de escalera de un vecindario está indignado porque hace 3 días que dio parte a la empresa de una avería y todavía no han acudido a arreglarla. ¿Cuál es la probabilidad de que se den este tipo de situaciones?

c) ¿Son X e Y independientes?

d) Calcúlese la probabilidad de que un ascensor haya funcionado con normalidad durante al menos 200 días y, una vez averiado, la empresa haya tardado medio día en ir a arreglarlo.

5.30. Sea (X, Y) una variable bidimensional continua con función de densidad:

$$f(x,y) = \begin{cases} \frac{2-x-y}{8} & -1 \leq x \leq 1, \quad -1 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Compruébese que X e Y son dependientes y que X^2 e Y^2 son independientes.

5.31. El portero de una comunidad de vecinos recoge la basura depositada por la familia del señor Pérez a la puerta de su casa en cualquier instante entre las 17 y las 19 horas. La variable aleatoria X, tiempo en minutos desde las 17 horas hasta la recogida, responde a la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{1}{120} & 0 < x < 120 \\ 0 & \text{en el resto} \end{cases}$$

Los señores Pérez son muy despistados y, en lugar de depositar la basura a las 17 horas (asegurando, así, que el portero se la lleve), lo hacen a partir de dicha hora en cualquier momento, Y, igualmente distribuido que X. ¿Qué porcentaje de veces quedará la basura en la puerta de la vivienda?

5.32. En la última reunión de accionistas de una red de supermercados se decidió realizar un estudio sobre la posible relación entre la proporción de clientes que compra sólo una vez al mes en estos comercios, X, y el beneficio mensual en millones de pesetas, Y.
Tras unos meses de trabajo estadístico, se obtuvo que la función de densidad conjunta de estas variables es:

$$f(x,y) = \begin{cases} k(x+y) \cdot e^{-y} & 0 < x < 1, \quad y > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese el valor de la constante k.
b) ¿Son X e Y independientes?
c) Calcúlese el porcentaje de meses en los que menos del 25 % de los clientes son de los que compran una vez al mes y los beneficios superen el millón de pesetas.
d) Si en un mes se han conseguido unos beneficios inferiores a dos millones de pesetas, ¿cuál es la probabilidad de que más de la mitad de los clientes pertenezca al grupo de los que compran una vez al mes?
e) Suponiendo que los beneficios de cada mes son independientes, obténgase la probabilidad de que la suma de los beneficios correspondientes a dos meses sea superior a 7 millones de pesetas.
f) ¿Cuál es la probabilidad de que en 6 meses con unos beneficios superiores a los tres millones de pesetas cada uno, el número de clientes que compran sólo una vez al mes sea, en todos ellos, inferior al 30 %?

5.33. El tiempo en minutos que un administrativo emplea en corregir un escrito es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} e^{-x} & x > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) ¿Cuál es la probabilidad de que tarde menos de 10 minutos en revisar dos escritos?

b) Si ha necesitado 15 minutos para corregir dos escritos, ¿cuál es la probabilidad de que empleara más de 10 minutos en revisar el primero?

5.34. Sean X_1 y X_2 variables aleatorias independientes tales que

$$p[X_i = 1] = p = 1 - p[X_i = 2] \text{ si } i = 1, 2 .$$

Sea Y_k la variable aleatoria que designa el número de variables X_i iguales a k con $k = 1, 2$. Obténgase la función de probabilidad de la variable (Y_1, Y_2).

5.35. La Asociación de Libreros de una ciudad ha publicado recientemente un trabajo sobre el tipo de literatura preferida por los lectores de edad comprendida entre 25 y 40 años.
Entre otros aspectos se analiza el número de libros de autores hispanos, X, que compra un individuo de un total, Y, de libros adquiridos a la semana, mediante la siguiente función de densidad conjunta:

$X \backslash Y$	0	1	2	3
0	2/8	2/8	0	0
1	0	1/8	1/8	0
2	0	0	1/8	0
3	0	0	0	1/8

a) Sabiendo que un individuo ha adquirido 2 libros en una semana, ¿cuál es la probabilidad de que ambos sean de autores hispanos?

b) Determínese el porcentaje de individuos que compran a la semana un libro de autor no hispano.

c) Si el número de libros que compra un individuo en una semana no depende de los que adquiere en cualquier otra, ¿cuál es la probabilidad de que el número de libros que un individuo adquiere al cabo de dos semanas, sea superior a 3?

5.36. En la siguiente tabla se da la distribución conjunta de las variables aleatorias X e Y:

$X \backslash Y$	1	2
0	0.05	0.15
1	0.3	0
2	0.05	0.45

a) Calcúlense las funciones de probabilidad y de distribución de las variables X e Y.

b) Obténgase la distribución de Y condicionada a que X tome el valor 2.

c) Hállense las siguientes probabilidades: $p[Y = 2 \mid X \leq 1]$, $p[X + Y > 2]$, $p[0 < X \leq 2 \mid Y = 1]$ y $p[X \cdot Y \leq 1]$.

d) ¿Son X e Y independientes?

e) Hállense las distribuciones conjuntas y marginales de $U = X + Y$ y $V = \max(X, Y)$.

5.37. Sea X la variable aleatoria que representa el peso, en toneladas, de una cierta mercancía almacenada a principios de mes de cara a satisfacer la demanda de los clientes. Sea Y el peso, en toneladas, de la mercancía vendida durante el mes. La función de densidad conjunta es:

$$f(x,y) = \begin{cases} \frac{1}{10x} & 0 < y < x < 10 \\ 0 & \text{en el resto} \end{cases}$$

a) Determínese la distribución de probabilidad del peso de la mercancía que queda almacenada a final de mes.

b) Hállese el porcentaje de meses en que la mercancía almacenada a primeros de mes sea superior a ocho toneladas y las ventas inferiores a cuatro toneladas.

c) Si un mes las ventas no superan las cinco toneladas, ¿cuál es la probabilidad de que al final de mes queden almacenadas más de tres toneladas?

5.38. Un grupo de empresas destina sus beneficios a inversión tecnológica, a inversión en cartera y a pago de dividendos entre sus accionistas, siendo X, Y y Z los porcentajes correspondientes a cada una de estas partidas. La función de densidad conjunta de X e Y viene dada por:

$$f(x,y) = \begin{cases} k & 0 \leq y \leq x\,, x+y \leq 50 \\ 0 & \text{en el resto} \end{cases}$$

a) ¿Cuánto vale k?

b) Hállense las funciones de densidad marginal de X y de Y.

c) Obténganse las funciones de densidad de Y condicionada a X y de X condicionada a Y.

d) Determínese la función de distribución y de densidad de Z.

5.39. Sea (X,Y) una variable aleatoria bidimensional con función de densidad:

$$f(x,y) = \begin{cases} 1 & 0 \leq x \leq 2, \quad \max(0,x-1) \leq y \leq \min(1,x) \\ 0 & \text{en el resto} \end{cases}$$

Hállese la función de densidad de la variable $U = X - Y$. ¿Son independientes las variables Y y U?

5.40. La compañía telefónica desea realizar un análisis sobre la repercusión que las nuevas tarifas han tenido en el número de llamadas. Teniendo en cuenta que las llamadas se clasifican en urbanas, interurbanas e internacionales, un estudio llevado a cabo sobre un grupo de familias, conduce a que las proporciones de llamadas urbanas, X, e interurbanas, Y, durante un mes tienen la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} 6x & x \geq 0\,,\ y \geq 0\,,\ x+y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Calcúlese la probabilidad de que el porcentaje de llamadas urbanas realizadas por una familia en un mes, sea superior al 70 %.

b) Hállese la probabilidad de que en una familia la proporción de llamadas urbanas de un cierto mes sea inferior a la de interurbanas.

c) Hállese la función de densidad correspondiente al porcentaje total de llamadas urbanas e interurbanas.

d) Determínese el porcentaje de meses en que la proporción de llamadas al extranjero para una cierta familia es superior a 0.2.

e) Si una familia no ha tenido conferencias con el extranjero, ¿cuál es la probabilidad de que al menos el 60 % de las llamadas hayan sido urbanas?

5.41. La variable aleatoria (X, Y) tiene la siguiente función de densidad:

$$f(x,y) = \begin{cases} kx \cdot y & x \geq 0, \quad y \geq 0, \quad x + y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) ¿Cuánto vale k?

b) Obténganse las funciones de densidad y de distribución marginales.

c) Hállese: $f_{(Y \mid X=x)}(y)$, $f_{(X \mid Y=y)}(x)$, $f_{(X \mid Y<1/2)}(x)$ y $f_{(Y \mid 0.5<X<0.7)}(y)$.

d) Calcúlense las siguientes probabilidades: $p[X \geq Y]$, $p[X \geq 1/2 \mid X + Y \leq 3/4]$ y $p[X^2 + Y^2 \leq 1]$.

e) ¿Son X e Y independientes?

f) Determínese un subconjunto $A \subset \mathbb{R}^2$ tal que $p(A) = 3/4$.

g) Hállese la distribución conjunta de $U = X + Y$ y $V = X - Y$.

h) Obténganse las funciones de distribución y de densidad de $Z = X \cdot Y$.

5.42. Compruébese que, dada una variable aleatoria bidimensional continua con función de densidad $f_{(X,Y)}(x, y)$ y una transformación (g, h) que verifique las hipótesis de la proposición 5.28, la función

$$f_{(U,V)}(u,v) = \begin{cases} f_{(X,Y)}(g^{-1}(u,v), h^{-1}(u,v)) \cdot |J| & \text{si } (u,v) \in D \\ 0 & \text{en el resto} \end{cases}$$

donde $D = (g, h)(\mathbb{R}^2)$, es realmente la función de densidad de una variable aleatoria bidimensional continua.

Capítulo 6

CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA UNIDIMENSIONAL

La siguiente situación servirá para ilustrar la necesidad de introducir una serie de nuevos conceptos.

Con el fin de llevar a cabo un plan de higiene y seguridad en el trabajo, una empresa contabiliza el número de accidentes laborales diarios. En el último mes, se han recogido los siguientes datos:

Número de accidentes diarios	Número de días del mes
0	10
1	12
2	5
3	2
4	1

Considerando la variable estadística X, *número de accidentes diarios*, la tabla siguiente recoge su distribución de frecuencias, donde n_i es la frecuencia absoluta y f_i la frecuencia relativa:

x_i	n_i	f_i
0	10	10/30
1	12	12/30
2	5	5/30
3	2	2/30
4	1	1/30
	$n = 30$	1

La tabla contiene toda la información sobre la variable estadística. Ahora bien, de cara a resumir ciertos aspectos de la distribución de frecuencias (promedio diario de accidentes, homogeneidad de los datos, forma de la distribución, etc.), que pongan *en claro* la información que contienen las observaciones y permitan la com-

paración con otras situaciones (número de accidentes diarios durante otro mes, número de accidentes en otra empresa, etc.), se calculan una serie de medidas características. Dichas medidas se clasifican según cuál sea el tipo de resumen que se pretende obtener en:

- Medidas de posición.
- Medidas de dispersión.
- Medidas de forma.

Así, entre las primeras, calcularíamos la media aritmética:

$$\overline{x} = \sum_{i=1}^{k} x_i \cdot f_i = 0 \cdot \frac{10}{30} + 1 \cdot \frac{12}{30} + 2 \cdot \frac{5}{30} + 3 \cdot \frac{2}{30} + 4 \cdot \frac{1}{30} = \frac{32}{30} = 1.07 \, .$$

Por tanto, el número medio diario de accidentes durante el mes estudiado, es de 1.07.

Otras medidas de posición, en concreto de posición central, que podemos calcular, son la moda (Mo) y la mediana (Me), que proporcionan el valor más frecuente y el valor central, respectivamente, de la variable.

Entre las medidas de dispersión podemos hallar la varianza,

$$\begin{aligned} S_X^2 &= \sum_{i=1}^{k} (x_i - \overline{x})^2 \cdot f_i = \\ &= \left(0 - \frac{32}{30}\right)^2 \cdot \frac{10}{30} + \left(1 - \frac{32}{30}\right)^2 \cdot \frac{12}{30} + \left(2 - \frac{32}{30}\right)^2 \cdot \frac{5}{30} + \\ &\quad + \left(3 - \frac{32}{30}\right)^2 \cdot \frac{2}{30} + \left(4 - \frac{32}{30}\right)^2 \cdot \frac{1}{30} = 1.06 \, , \end{aligned}$$

y su raíz cuadrada positiva, la desviación típica,

$$S_X = +\sqrt{S_X^2} = 1.03 \, .$$

Estas dos medidas nos dan una idea de la dispersión de los valores de la variable alrededor de su media y, por tanto, del grado de representatividad de ésta como resumen global de la distribución de frecuencias. Si quisiéramos relativizarlas teniendo en cuenta el tamaño de la variable, calcularíamos el coeficiente de variación:

$$CV = \frac{S_X}{\overline{X}} = 0.97 \, .$$

Igualmente, podríamos calcular otras medidas que proporcionen una idea de la posición de la variable (cuantiles), de su dispersión (recorrido intercuartílico, desviación absoluta media, etc.) o del grado de simetría y aplastamiento de la distribución de frecuencias.

Todas estas medidas tienen su fundamento en las frecuencias observadas para los distintos valores de la variable, es decir, son **medidas empíricas**.

Después de observar el comportamiento del número diario de accidentes durante distintos meses, un estadístico ha decidido asignar a dicha variable una distribución de probabilidad que trate de explicar su comportamiento futuro. En el siguiente cuadro se recoge dicha distribución de probabilidad:

x_i	$p[X = x_i]$
0	0.30
1	0.35
2	0.20
3	0.08
4	0.05
5	0.02

De cara a resumir los aspectos más relevantes de esta distribución, vamos a construir una serie de **medidas teóricas** que nos sitúen la variable en la recta real (medidas de posición), que nos proporcionen una idea del grado de concentración de los valores alrededor de un valor central (medidas de dispersión) o que nos informen sobre el perfil de la función de probabilidad o de densidad (medidas de forma).

De nuevo, hablaremos de media (o esperanza), moda, mediana, varianza o coeficientes de asimetría y apuntamiento, siendo la construcción de todas estas medidas similar al de las medidas descriptivas antes citadas. Si la distribución de probabilidad es dicreta, sólo tendremos que cambiar frecuencia por probabilidad, en conceptos y expresiones; si es continua, este cambio deberá ir acompañado por un salto matemático de lo discreto a lo continuo que, en la mayoría de las ocasiones, se traducirá en transformar sumas en integrales.

6.1 Esperanza matemática

La media aritmética es el principal resumen de una variable estadística. Bajo ciertas condiciones (dispersión pequeña), la media es un buen representante de toda la distribución de frecuencias. Su interpretación es clara: por un lado, es centro de gravedad de la distribución; por otro, es el valor de reparto igualitario de la suma total de los valores de la variable.

En este apartado desarrollaremos la media referida a una distribución de probabilidades: la esperanza matemática de una variable (también hablaremos de valor medio, media o valor esperado).

El concepto de esperanza tiene sus orígenes históricos en la ganancia que un jugador *esperaba* obtener tras sucesivas partidas de un juego de azar.

Ejemplo 6.1. Se lanza un dado, si el resultado obtenido es *impar*, el jugador recibirá de la banca tantas pesetas como indique el dado. Si sale *par*, el jugador abonará a la banca dicha cantidad.

Se realiza un número grande, n, de jugadas. Suponiendo que el dado no está cargado,

cada una de las caras aparecerá $n/6$ veces aproximadamente. La ganancia del jugador, después de esas n jugadas, será aproximadamente

$$1 \cdot \frac{n}{6} + (-2) \cdot \frac{n}{6} + 3 \cdot \frac{n}{6} + (-4) \cdot \frac{n}{6} + 5 \cdot \frac{n}{6} + (-6) \cdot \frac{n}{6} = -\frac{3n}{6} ,$$

es decir, ganará una peseta en las $n/6$ jugadas donde salga la cara con el uno, perderá dos pesetas en las $n/6$ jugadas en que salga la cara con el dos, etc.

Por tanto, la ganancia media por jugada será la n-ésima parte de la cantidad anterior:

$$\text{Ganancia esperada} = 1 \cdot \frac{1}{6} + (-2) \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + (-4) \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + (-6) \cdot \frac{1}{6} = -\frac{3}{6} = -0.5 .$$

Luego, el jugador espera obtener -0.5 pesetas por jugada, es decir, espera perder dicha cantidad: el juego no le es favorable.

Construyendo una variable X, *ganancia del jugador*, ésta tomará los siguientes valores con su probabilidad correspondiente:

$$X: \left| \begin{array}{ll} 1 & p[X=1] = 1/6 \\ -2 & p[X=-2] = 1/6 \\ 3 & p[X=3] = 1/6 \\ -4 & p[X=-4] = 1/6 \\ 5 & p[X=5] = 1/6 \\ -6 & p[X=-6] = 1/6 \end{array} \right.$$

La ganancia esperada, esperanza de X, que designaremos por $E(X)$, será la suma de los valores de la variable multiplicados por su probabilidad. Así,

$$E(X) = 1 \cdot \frac{1}{6} + (-2) \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + (-4) \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + (-6) \cdot \frac{1}{6} = -0.5 \text{ pesetas.} \quad \triangleleft$$

El ejemplo anterior nos conduce a la definición de esperanza de una variable aleatoria.

Definición 6.1 *Sea X una variable aleatoria discreta que toma los valores $x_1, \ldots, x_n, \ldots$ y cuya función de probabilidad es f_X. Se llama esperanza de X al número*

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot f_X(x_i) = \sum_{i=1}^{\infty} x_i \cdot p[X = x_i] .$$

Dicho número existe siempre que $\sum_{i=1}^{\infty} |x_i| \cdot f_X(x_i) < \infty$, es decir, siempre que la serie sea absolutamente convergente.

Si la variable discreta fuese finita, la definición anterior sería igualmente válida, considerando tan sólo un número finito de sumandos (en este caso no hay problemas de existencia de $E(X)$).

Ejemplo 6.2. Una rifa consta de 100 papeletas que se venden a un precio de 10 pesetas, habiendo un único premio de 500 pesetas. Un jugador compra tres boletos, ¿cuál es la ganancia que espera obtener?

En el caso de que no consiga premio, el jugador habrá perdido las 30 pesetas que invirtió; por el contrario, si obtuviese el premio, ganaría 470 pesetas. Hallemos la probabilidad de no obtener premio,

$$\begin{aligned} p(\text{ no obtener premio }) &= p(\overline{P_1}\ \overline{P_2}\ \overline{P_3}) = p(\overline{P_1}) \cdot p(\overline{P_2} \mid \overline{P_1}) \cdot p(\overline{P_3} \mid \overline{P_1}\ \overline{P_2}) = \\ &= \frac{99}{100} \cdot \frac{98}{99} \cdot \frac{97}{98} = \frac{97}{100}\,, \end{aligned}$$

donde P_i designa el suceso *obtener premio con la i-ésima papeleta.* Resumiendo toda la información disponible sobre la variable aleatoria, X, *ganancia del jugador*, tendremos:

$$X: \left| \begin{array}{ll} -30 & p[X=-30] = p(\text{no obtener premio}) = 97/100 \\ 470 & p[X=470] = p(\text{obtener premio}) = 3/100 \end{array} \right.$$

De aquí, concluimos que la ganancia esperada es

$$E(X) = -30 \cdot \frac{97}{100} + 470 \cdot \frac{3}{100} = -15 \text{ pesetas.}$$

Esto indica que, si el jugador jugase indefinidamente en esta rifa, siempre comprando tres papeletas, obtendría una ganancia (pérdida, en este caso) media por jugada de -15 pesetas. Obsérvese, por otro lado, que la esperanza de una variable no tiene por qué ser un valor de la misma. ◁

La definición anterior se puede reinterpretar para variables continuas de la siguiente manera:

Definición 6.2 *Sea X una variable aleatoria continua con función de densidad f_X. Se llama esperanza de X al número*

$$E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)dx\,.$$

Dicho número existe siempre que $\int_{-\infty}^{\infty} |x| \cdot f_X(x)dx < \infty$, es decir, siempre que la integral sea absolutamente convergente.

Ejemplo 6.3. El tiempo de vida, en años, de una cierta componente de una máquina se modeliza mediante la siguiente función de densidad:

$$f_X(x) = \left\{ \begin{array}{ll} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{array} \right.$$

El tiempo medio de vida de dicha componente es

$$E(X) = \int_0^{\infty} x \cdot e^{-x}dx = 1 \text{ año,}$$

integral que se ha calculado *por partes*, resolviendo la indeterminación $\infty \cdot 0$ mediante la regla de L'HÔPITAL. Por tanto, se espera que dicha componente tenga una duración de un año. ◁

Si la variable X fuese mixta (con una parte discreta, $\mathcal{D}$, y otra continua, $\mathcal{C}$), podríamos definir de igual forma su esperanza, sumando para los valores discretos e integrando para los continuos; es decir, si f_X es su función de probabilidad/densidad,

$$E(X) = \sum_{x_i \in \mathcal{D}} x_i \cdot f_X(x_i) + \int_{\mathcal{C}} x \cdot f_X(x) dx \, .$$

Ejemplo 6.4. Consideremos la variable mixta descrita en el ejemplo 4.11. Recordemos que dicha variable, *número de litros por* m^2 *diarios* recogidos en un observatorio meteorológico, se modelizaba según la siguiente función de distribución:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - 0.5 \cdot e^{-x} & x \geq 0 \end{cases}$$

Como vimos, X es discreta en el cero (la función de distribución presenta un salto, una discontinuidad), siendo 0.5 la probabilidad de dicho punto; por el contrario, en el intervalo $(0, \infty)$ es continua, siendo su función de densidad $0.5 \cdot e^{-x}$. Entonces, la cantidad esperada de lluvia será:

$$E(X) = 0 \cdot p[X = 0] + \int_0^{\infty} x \cdot 0.5 \cdot e^{-x} dx = 0.5 \text{ litros},$$

es decir, se espera que llueva diariamente medio litro por m^2. ◁

La esperanza de una variable aleatoria, que normalmente designaremos por μ, constituye un resumen adecuado de su distribución de probabilidad, si se cumplen dos condiciones. En primer lugar, que no existan *valores raros* de la variable, es decir, valores notoriamente distantes de la zona en que la variable aleatoria se encuentra mayoritariamente. Si no existen dichos valores raros, la esperanza será un resumen aceptable de la variable aleatoria, entendida, al menos, como un valor central de la misma.

La segunda condición se refiere al grado de representatividad de la esperanza; este grado de representatividad tiene que ver con la dispersión de los valores de la variable, como analizaremos en la siguiente sección. Si la variable está muy dispersa, su esperanza, aun siendo un valor central, no será un valor recomendable como resumen de la distribución de probabilidad.

Veamos dos ejemplos que aclaran el primer comentario. El segundo lo ilustraremos con el ejemplo 6.9.

Ejemplo 6.5. La construcción de la urbanización La Moralina, con chalets de alto standing, en un barrio de chabolas, hace que convivan, en esa zona de la ciudad, familias con muy distintos niveles sociales, económicos y culturales.

La concejalía de asuntos sociales ha estudiado el *número de hijos por familia*, X, obteniendo la distribución de probabilidad siguiente:

$$p[X = 0] = 0.4, \quad p[X = 1] = 0.3, \quad p[X = 2] = 0.2 \quad \text{y} \quad p[X = 23] = 0.1 \, .$$

El número esperado de hijos es 3, valor que representa escasamente a una distribución que asigna una probabilidad de 0.9 al suceso *tener a los sumo dos hijos.* ◁

El siguiente ejemplo incide en la misma idea, desde otro punto de vista; en concreto, muestra que, si introducimos un valor suficientemente raro en una distribución, aunque éste tenga una probabilidad muy pequeña, podremos obtener cualquier valor como esperanza de la variable.

Ejemplo 6.6. Sea X una variable aleatoria tal que

$$p[X = 1] = 0.99 \quad \text{y} \quad p[X = a] = 0.01 \, .$$

Determinemos el valor de a de forma que la esperanza de X sea un valor, m, cualquiera. Evidentemente,

$$E(X) = 1 \cdot 0.99 + a \cdot 0.01 \, ,$$

con lo cual, el valor de a se calcula resolviendo la ecuación

$$0.99 + 0.01 \cdot a = m \, .$$

De donde,

$$a = \frac{m - 0.99}{0.01} = 100(m - 0.99) \, . \triangleleft$$

Las definiciones de esperanza matemática de una variable aleatoria X, se pueden generalizar para funciones $g(X)$ que sean también variables aleatorias.

Definición 6.3 *Sea X una variable aleatoria discreta con función de probabilidad f_X. Sea $g(X)$ una variable aleatoria función de X. Se llama esperanza de $g(X)$ al número*

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i) \cdot f_X(x_i) = \sum_{i=1}^{\infty} g(x_i) \cdot p[X = x_i] \, .$$

Esta definición sólo tiene sentido cuando $\sum_{i=1}^{\infty} |g(x_i)| \cdot f_X(x_i) < \infty$.

Definición 6.4 *Sea X una variable aleatoria continua con función de densidad f_X. Sea $g(X)$ una variable aleatoria continua función de X. Se llama esperanza de $g(X)$ al número*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx \, .$$

En este caso, existe la esperanza siempre y cuando $\int_{-\infty}^{\infty} |g(x)| \cdot f_X(x) dx < \infty$.

Siendo X continua, pudiera ocurrir que $g(X)$ fuese una variable discreta. En ese caso, hallaríamos las probabilidades de cada valor de $g(X)$ integrando la función de densidad f_X en distintos intervalos.

Las definiciones 6.3 y 6.4 pueden inducir a confusión a algún lector; en efecto, si X es una variable aleatoria y $g(X)$ es una función suya que, a su vez, es variable aleatoria, entonces, con las técnicas estudiadas en la sección 4.8, podríamos deducir la función de probabilidad o de densidad de la variable transformada, $Y = g(X)$, a partir de la de X, calculando así su esperanza, con arreglo a las definiciones 6.1 o 6.2.

En ese caso, ¿qué sentido tienen las definiciones 6.3 y 6.4? Además de confirmar, implícitamente, que las dos técnicas de cálculo de $E[g(X)]$ conducen al mismo resultado, sugieren que el procedimiento que con ellas se propone es, en general, el más conveniente para el cálculo de la esperanza de una función de X. De hecho, habitualmente se manejan varias funciones de una variable aleatoria, por lo que el método de obtención de la esperanza de la variable transformada, calculando previamente su distribución, nos llevaría a la realización de múltiples cambios de variable.

Ejemplo 6.7. Para el mantenimiento de la componente, cuyo tiempo de vida habíamos descrito en el ejemplo 6.3, se suscribe una póliza. De acuerdo con ella, el coste de dicho mantenimiento, en millones de pesetas, Y, será una función del tiempo de vida de la componente, X; en concreto,

$$Y = g(X) = 2X^2 + 1 \, .$$

Queremos calcular la cantidad que espera abonar la empresa en concepto de mantenimiento, esto es,

$$E(Y) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx = \int_0^{\infty} (2x^2 + 1) \cdot e^{-x} dx = 5 \text{ millones de pesetas.}$$

El lector que desee resolver la integral anterior deberá aplicar dos veces el método de integración por partes.

Otra forma de actuar consiste en hallar la esperanza de la nueva variable Y, obteniendo previamente su función de densidad. Así, para $y \geq 1$,

$$F_Y(y) = p[Y \leq y] = p[2X^2 + 1 \leq y] = p\left[X \leq \sqrt{\frac{y-1}{2}}\right] = F_X\left(\sqrt{\frac{y-1}{2}}\right) ,$$

de donde, derivando, tendremos

$$f_Y(y) = f_X(\sqrt{(y-1)/2}) \cdot \frac{1}{\sqrt{8(y-1)}} = \frac{e^{-\sqrt{(y-1)/2}}}{\sqrt{8(y-1)}} ,$$

función que se anula para $y < 1$. En consecuencia,

$$E(Y) = \int_1^{\infty} y \cdot \frac{e^{-\sqrt{(y-1)/2}}}{\sqrt{8(y-1)}} dy = 5 \text{ millones de pesetas.}$$

Otra empresa propone un contrato de mantenimiento que obliga a pagar dos millones de pesetas por cada año, o parte de año, en que esté en funcionamiento dicha componente. ¿Cuál de las dos proposiciones resulta más ventajosa?

El coste del nuevo contrato, Z, será de 2 millones, si la componente no dura más de un año, de 4 millones, si su duración es superior al año e inferior a los dos años, de 6 millones, si supera los dos años pero no el tercero, y así sucesivamente. Entonces, la variable es

$$Z : \left| \begin{array}{ll} 2 & X \leq 1 \\ 4 & 1 < X \leq 2 \\ \vdots & \vdots \\ 2k & k-1 < X \leq k \\ \vdots & \vdots \end{array} \right.$$

Por tanto,

$$p[Z = 2k] = p[k-1 < X \leq k] = \int_{k-1}^{k} e^{-x}dx = e^{-(k-1)} - e^{-k} ,$$

y la esperanza,

$$\begin{aligned} E(Z) &= 2 \cdot (1 - e^{-1}) + 4 \cdot (e^{-1} - e^{-2}) + \cdots + 2k \cdot (e^{-(k-1)} - e^{-k}) + \cdots = \\ &= \frac{2}{1 - e^{-1}} = 3.16 \text{ millones de pesetas,} \end{aligned}$$

valor obtenido sumando una progresión geométrica ilimitada de razón e^{-1} y primer término 2.

De todo lo anterior, deducimos que la segunda proposición de mantenimiento es más ventajosa, pues su coste esperado supera, en poco, los tres millones, frente a los cinco de la primera opción. Esto no quiere decir que la segunda opción sea siempre más ventajosa. Si, por ejemplo, la componente sólo durase medio año, el coste de su mantenimiento sería de 1.5 millones, en caso de haber suscrito la primera póliza, y de dos millones, en el otro caso. Ahora bien, teniendo en cuenta el comportamiento aleatorio del tiempo de vida de la componente, es preferible la segunda póliza, porque su coste esperado es inferior. ◁

En las distintas definiciones de esperanza, se ha hecho hincapié en la necesidad de exigir que la esperanza del valor absoluto de la variable sea finita; esto es, $E(|X|) < \infty$. En la mayoría de las ocasiones (prácticamente todas las que veremos en este libro), esta condición se verifica, por lo cual obviaremos su comprobación. A continuación, veremos un ejemplo de una variable continua para la cual no existe la esperanza matemática. Esta situación podría ocurrir, también, en ciertas variables discretas infinitas.

Ejemplo 6.8. Sea X una variable aleatoria cuya función de densidad es

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} \qquad -\infty < x < \infty .$$

Queremos calcular el valor esperado de X, por lo que, previamente, nos planteamos el problema de la existencia de dicho valor:

$$\begin{aligned} E(|X|) &= \int_{-\infty}^{\infty} |x| \cdot f(x)dx = \\ &= \int_{-\infty}^{0} -x \cdot \frac{1}{\pi} \cdot \frac{1}{1 + x^2}dx + \int_{0}^{\infty} x \cdot \frac{1}{\pi} \cdot \frac{1}{1 + x^2}dx = \\ &= \frac{1}{2\pi} \left\{ [-\ln(1 + x^2)]_{x=-\infty}^{0} + [\ln(1 + x^2)]_{x=0}^{\infty} \right\} , \end{aligned}$$

integral que no es convergente, es decir, no toma un valor finito, ya que $\ln(1 + x^2)$ tiende a ∞ cuando x tiende a ∞ o a $-\infty$. Por tanto, no existe $E(X)$. El lector puede comprobar que su cálculo nos conduciría a una indeterminación.

La idea de que la esperanza es el centro de gravedad de la distribución de probabilidad, podría conducir a la afirmación errónea de que $E(X)$ vale, en este caso, cero (téngase en cuenta que la función de densidad es simétrica respecto del cero, siendo este punto el

inequívoco centro de gravedad). Ahora bien, obrar de esta forma, obviando las formalizaciones matemáticas, llevaría a la larga a resultados, no sólo inciertos, sino contrarios a la misma intuición. ◁

6.1.1 Propiedades de la esperanza

P1. Si X es una variable aleatoria acotada, esto es, si para una cierta constante M, $|X| < M$, entonces, existe la esperanza de X.

Demostración. Tenemos que probar que $E(|X|) < \infty$.
Si X es discreta,

$$E(|X|) = \sum_{i=1}^{\infty} |x_i| \cdot p[X = x_i] < \sum_{i=1}^{\infty} M \cdot p[X = x_i] = M \, .$$

Si X es continua,

$$E(|X|) = \int_{-\infty}^{\infty} |x| \cdot f(x)dx < \int_{-\infty}^{\infty} M \cdot f(x)dx = M \, .$$

Como, en ambos casos, $E(|X|) < M$, aseguramos la existencia de $E(X)$. ♣

La propiedad anterior se verifica, también, si $p(|X| < M) = 1$, esto es, si el conjunto de valores para los que X es mayor que cierta constante, tiene probabilidad nula.

P2. Si X es una variable aleatoria acotada superior e inferiormente, es decir, si existen dos constantes a y b, tales que $a \leq X \leq b$, la esperanza de X está comprendida entre a y b, esto es,

$$a \leq E(X) \leq b \, .$$

Demostración. La propiedad anterior nos asegura la existencia de la esperanza. Entonces, y suponiendo X discreta, para todo i se verifica que $a \leq x_i \leq b$. Por tanto,

$$a \cdot p[X = x_i] \leq x_i \cdot p[X = x_i] \leq b \cdot p[X = x_i] \, .$$

Luego,

$$\sum_{i=1}^{\infty} a \cdot p[X = x_i] \leq \sum_{i=1}^{\infty} x_i \cdot p[X = x_i] \leq \sum_{i=1}^{\infty} b \cdot p[X = x_i] \, ,$$

con lo que

$$a \leq E(X) \leq b \, .$$

De idéntica forma razonaríamos en el caso continuo. ♣

Asimismo, esta propiedad se verifica si $p[a \leq X \leq b] = 1$, esto es, si X está comprendido entre a y b, salvo en un conjunto de probabilidad nula.

P3. Si X es una variable aleatoria no negativa, es decir, si $X \geq 0$, y existe su esperanza, dicha esperanza también será no negativa.

El lector puede demostrar esta propiedad, que del mismo modo se cumple si $p[X \geq 0] = 1$, como caso particular de la propiedad anterior.

P4. La esperanza de una constante es la propia constante.

Demostración. Una constante es una variable aleatoria degenerada, es decir, una variable que toma un único valor K, con probabilidad uno, $p[X = K] = 1$. Por tanto,

$$E(X) = K \cdot p[X = K] = K \,. \clubsuit$$

P5. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ funciones de X que, a su vez, son variables aleatorias, existiendo sus respectivas esperanzas. Sean a y b constantes cualesquiera. Entonces,

$$E[a \cdot g(X) + b \cdot h(X)] = a \cdot E[g(X)] + b \cdot E[h(X)] \,.$$

Demostración. Si las variables aleatorias son continuas, y f_X es la función de densidad de X, entonces,

$$\begin{aligned} E[a \cdot g(X) + b \cdot h(X)] &= \int_{-\infty}^{\infty} [a \cdot g(x) + b \cdot h(x)] \cdot f_X(x)dx = \\ &= a \cdot \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx + b \cdot \int_{-\infty}^{\infty} h(x) \cdot f_X(x)dx = \\ &= a \cdot E[g(X)] + b \cdot E[h(X)] \,. \end{aligned}$$

Mediante un razonamiento análogo, demostraríamos la propiedad para variables aleatorias discretas. $\clubsuit$

De esta propiedad, que se conoce como propiedad de linealidad de la esperanza, deducimos, eligiendo convenientemente las constantes a y b y las funciones g y h, los siguientes corolarios:

a) $E[g(X) + h(x)] = E[g(X)] + E[h(X)]$.
b) $E[a \cdot g(X) + b] = a \cdot E[g(X)] + b$.
c) $E(a \cdot X + b) = a \cdot E(X) + b$.
d) $E[a \cdot g(X)] = a \cdot E[g(X)]$.
e) $E(a \cdot X) = a \cdot E(X)$.

El tercer corolario proporciona la esperanza de una importante transformación de X, $Y = a \cdot X + b$. Un caso particular de la misma es lo que se conoce como **cambio de origen y de escala**.

Sea X una variable aleatoria. Decimos que *la variable Y es la transformada de X por un cambio de origen y de escala, si*

$$Y = \frac{X - o}{e},$$

donde o y e $(e > 0)$ son dos constantes cualesquiera. Resulta evidente que esta transformación es un caso particular de la transformación $Y = a \cdot X + b$ para $a = \frac{1}{e}$ y $b = -\frac{o}{e}$.

El cambio de origen y de escala puede entenderse como un proceso en dos etapas. En la primera, denominada cambio de origen, la variable X se transforma en $U = X - o$; mediante esta transformación, el conjunto de valores de la variable se *mueve*, pero manteniéndose la posición relativa y la distancia entre los mismos.

En la segunda etapa, la variable, ya cambiada de origen, $U = X - o$, se transforma en $Y = \frac{X-o}{e}$, es decir, cambia de escala. Si e es mayor que uno, se *acercan* los valores de la variable unos a otros, esto es, se concentran; si e es menor que uno, los valores se *alejan*, aumentan sus distancias, es decir, se dispersan.

Aplicando el segundo corolario demostrado anteriormente, podemos comprobar el efecto que produce el cambio de origen y de escala en la esperanza de la variable. Así,

$$E(Y) = E\left(\frac{X - o}{e}\right) = \frac{E(X) - o}{e},$$

o, lo que es lo mismo,

$$\mu_Y = \frac{\mu_X - o}{e}.$$

La figura 6.1 recoge la transformación producida por el cambio de origen sobre la función de densidad de una distribución continua.

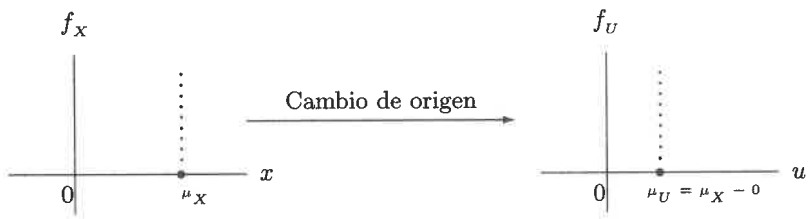


FIG. 6.1.

Por su parte, una vez realizado el cambio de origen, la figura 6.2 describe la variación de la función de densidad debida a un cambio de escala.

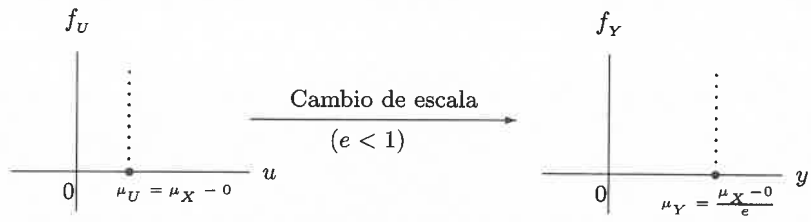


FIG. 6.2.

Si la variable fuese discreta, el resultado de la transformación sería similar.

Obviamente, si $e = 1$, tendríamos solamente un cambio en el origen, y si $o = 0$, un cambio de escala. Por otro lado, aunque desde el punto de vista formal no habría inconveniente en considerar valores negativos de e, dicha consideración carece de sentido desde el punto de vista del cambio de escala, por lo que no parece aconsejable utilizar dicha denominación en ese caso.

Ejemplo 6.9. Sea X una variable discreta con la siguiente función de probabilidad:

X	$p[X = x]$
1	0.2
2	0.4
3	0.3
5	0.1

La esperanza de X será:

$$E(X) = 1 \cdot 0.2 + 2 \cdot 0.4 + 3 \cdot 0.3 + 5 \cdot 0.1 = 2.4 .$$

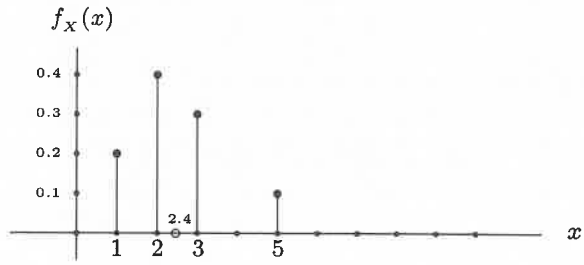


FIG. 6.3.

Transformemos la variable X, cuya función de probabilidad se representa en la figura 6.3, mediante un cambio de origen y de escala. Sea Y la nueva variable

$$Y = \frac{X + 5}{2} .$$

Consideremos en primer lugar, únicamente, el cambio de origen, es decir, $U = X + 5$. Su distribución será:

U	$p[U = u]$
6	0.2
7	0.4
8	0.3
10	0.1

La esperanza de U, cuya función de probabilidad se representa en la figura 6.4, es:

$$E(U) = E(X) + 5 = 7.4 \, .$$

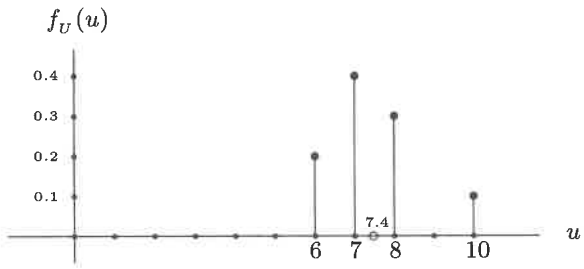


FIG. 6.4.

Como se observa, la variable se ha movido de posición, pero la distancia entre sus valores se mantiene.

A continuación, consideraremos el cambio completo, esto es,

$$Y = \frac{U}{2} = \frac{X + 5}{2} \, .$$

La función de probabilidad de Y, cuya esperanza será

$$E(Y) = \frac{E(U)}{2} = \frac{E(X) + 5}{2} = 3.7 \, ,$$

es:

Y	$p[Y = y]$
3	0.2
3.5	0.4
4	0.3
5	0.1

Su gráfica aparece en la figura 6.5.

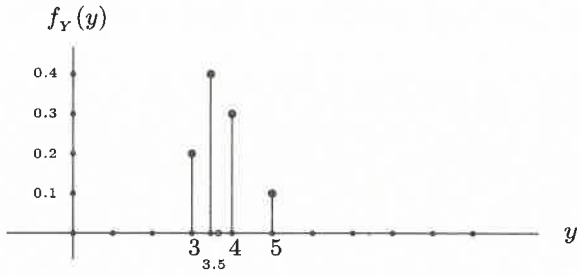


FIG. 6.5.

Advierta el lector que la posición global de la variable ha cambiado y, además, los valores de la misma se han concentrado (téngase en cuenta que $e = 2$, es decir, hemos reducido la escala de la variable). ◁

P6. Sea X una variable aleatoria y sean $g(X)$ y $h(X)$ dos funciones de X, variables aleatorias cuyas esperanzas existen. Si $g(X) \leq h(X)$, entonces,

$$E[g(X)] \leq E[h(X)] .$$

Demostración. Como $g(X) \leq h(X)$, la variable $h(X) - g(X)$ es no negativa y, por tanto, teniendo en cuenta la tercera propiedad,

$$E[h(X) - g(X)] \geq 0 .$$

De aquí, por la propiedad de linealidad,

$$E[g(X)] \leq E[h(X)] . \clubsuit$$

La propiedad será igualmente válida cuando $p[g(X) \leq h(X)] = 1$.

P7. Si existe la esperanza de una variable aleatoria X, entonces,

$$|E(X)| \leq E(|X|) .$$

Demostración. Si X es una variable aleatoria continua,

$$|E(X)| = \left| \int_{-\infty}^{\infty} x \cdot f(x)dx \right| \leq \int_{-\infty}^{\infty} |x| \cdot f(x)dx = E(|X|) .$$

De igual forma, teniendo en cuenta que el valor absoluto de una suma es menor o igual que la suma de los valores absolutos, demostraríamos esta propiedad para variables discretas. ♣

P8. Si X es una variable aleatoria cuya función de probabilidad o de densidad es simétrica respecto de un valor a, la esperanza, si existe, es dicho valor a.

Demostración. Supongamos que X es continua y que su función de densidad presenta simetría respecto del cero, es decir, $f_X(x) = f_X(-x)$. Entonces,

$$E(X) = \int_{x=-\infty}^{\infty} x \cdot f_X(x)dx = \int_{x=-\infty}^{0} x \cdot f_X(x)dx + \int_{x=0}^{\infty} x \cdot f_X(x)dx \, .$$

Haciendo el cambio de variable $x = -t$ en la primera integral, resulta:

$$\int_{t=\infty}^{0} -t \cdot f_X(-t)(-dt) + \int_{x=0}^{\infty} x \cdot f_X(x)dx = -\int_{t=0}^{\infty} t \cdot f_X(t)dt + \int_{x=0}^{\infty} x \cdot f_X(x)dx \, .$$

Por tanto, dado que ambas integrales coinciden, la esperanza de X es cero.

Si la función de densidad de X es simétrica respecto de a, es decir, $f_X(a+x) = f_X(a-x)$, entonces, la de la variable $Y = X - a$ es simétrica respecto de cero, con lo cual,

$$E(Y) = E(X - a) = 0 \, .$$

Deducimos, así, que $E(X) = a$.

De la misma manera, se demostraría el resultado para variables discretas. ♣

Si $a = 0$, esto es, si la función de probabilidad o de densidad es simétrica respecto del eje de ordenadas, la propiedad anterior indica que la esperanza de X es igual a 0. Veremos una generalización de este resultado cuando estudiemos los llamados momentos no centrales.

Por abuso de lenguaje, hablaremos de variable aleatoria simétrica, o bien de distribución simétrica, cuando lo sea su función de probabilidad o de densidad.

6.2 Momentos de una variable aleatoria

Dada una variable aleatoria, X, interesa considerar la esperanza de ciertas funciones $g(X)$ que resuman algunos aspectos de la distribución de probabilidad de X.

Si $g(X) = X^k$, donde $k = 0, 1, \ldots$, definimos

$$\alpha_k = E(X^k) \, ,$$

que llamaremos **momento centrado respecto del origen** de orden k o **momento no central**.

Si $g(X) = (X - C)^k$, para $k = 0, 1, \ldots$, tenemos

$$E(X - C)^k \, ,$$

que denominaremos **momento centrado respecto de** C de orden k.

De especial interés es el caso en que C es la esperanza de X, μ:

$$\mu_k = E[(X - E(X))^k] = E[(X - \mu)^k]$$

es el **momento centrado respecto de la media** de orden k o **momento central**.

Si $g(X) = X(X-1)\cdots(X-k+1)$, tendremos

$$E[X(X-1)\cdots(X-k+1)]\,,$$

denominado **momento factorial** de orden k.

Considerando las funciones anteriores en valor absoluto, se obtienen los correspondientes **momentos absolutos**.

Todos los momentos, como esperanzas que son, presentan el problema de su existencia en los términos planteados en la sección anterior. Ahora bien, podemos asegurar que la existencia de un momento de un cierto orden implica la existencia de los momentos de orden inferior.

Proposición 6.1 *Dada una variable aleatoria X, si existe un momento cualquiera de orden t, existen todos los momentos de orden inferior.*

Demostración. Realicemos la demostración para los momentos no factoriales, utilizando la notación común $g(X) = [h(X)]^k$, donde $h(X)$ se concretará de la siguiente forma:

- $h(X) = X$ para los momentos respecto del origen,
- $h(X) = X - C$ para los momentos respecto de C,
- $h(X) = X - E(X)$ para los momentos respecto de la media,
- $h(X) = |X|$, $h(X) = |X - C|$ y $h(X) = |X - E(X)|$ para los distintos momentos absolutos.

Entonces, probaremos que, si existe $E[(h(X))^t]$, existe $E[(h(X))^s]$, para todo valor de s inferior a t.

Para una variable aleatoria continua X, con función de densidad f_X,

$$\begin{aligned}
E[|h(X)|^s] &= \int_{x=-\infty}^{\infty} |h(x)|^s \cdot f_X(x)dx = \\
&= \int_{x/|h(x)|\leq 1} |h(x)|^s \cdot f_X(x)dx + \int_{x/|h(x)|>1} |h(x)|^s \cdot f_X(x)dx \leq \\
&\leq \int_{x/|h(x)|\leq 1} 1 \cdot f_X(x)dx + \int_{x/|h(x)|>1} |h(x)|^t \cdot f_X(x)dx \leq \\
&\leq \int_{x=-\infty}^{\infty} f_X(x)dx + \int_{x=-\infty}^{\infty} |h(x)|^t \cdot f_X(x)dx = \\
&= 1 + E[|h(X)|^t] < \infty.
\end{aligned}$$

Existirá, por tanto, $E[(h(X))^s]$.

Si X fuera discreta, la demostración sería análoga. ♣

En realidad, en la proposición anterior hemos demostrado que, si existe un momento de orden t, existe cualquier momento de orden inferior del mismo tipo; por ejemplo, si existiese μ_8, existirían μ_7, μ_6, ... Ahora bien, también está implícito el hecho de que existen todos los momentos de cualquier otro tipo de ese orden y de

órdenes inferiores; así, si existe μ_8, existirán, también, α_8, α_7, ..., al igual que los correspondientes momentos absolutos. Esto es debido a que, como comprobaremos en la proposición 6.2, los momentos centrales pueden expresarse como combinación lineal de los momentos no centrales de orden igual e inferior, y viceversa. En cuanto a los momentos absolutos, se exige su existencia en la definición del resto de los momentos; por ejemplo, $E[(X-\mu)^3]$ existe si, y sólo si, existe $E[|(X-\mu)^3|]$.

La proposición es igualmente válida para los momentos factoriales, dado que éstos se pueden poner como combinación lineal de los momentos respecto del origen de orden igual e inferior al dado. Así,

$$E[X(X-1)\cdots(X-k+1)] = E(X^k) + b_{k-1}E(X^{k-1}) + \cdots + b_1 E(X) .$$

De entre todos los momentos considerados, son de especial interés los momentos centrales y no centrales y, en concreto, los de ciertos órdenes.

6.2.1 Momentos centrales y no centrales

Como hemos definido anteriormente, *el momento centrado respecto del origen de orden k ($k = 0, 1, \ldots$) o momento no central de una variable aleatoria X, es el valor, si existe,*

$$\alpha_k = E(X^k) ,$$

es decir, si X es discreta,

$$\alpha_k = \sum_{i=1}^{\infty} x_i^k \cdot p[X = x_i]$$

y, si es continua,

$$\alpha_k = \int_{-\infty}^{\infty} x^k \cdot f_X(x)dx .$$

Para $k = 0$, dicho momento siempre vale la unidad:

$$\alpha_0 = E(X^0) = E(1) = 1 .$$

Para $k = 1$, el momento respecto del origen es la esperanza de la variable:

$$\alpha_1 = E(X) ,$$

que, como ya hemos dicho, designaremos normalmente por μ o μ_X.

Para $k = 2$,

$$\alpha_2 = E(X^2) ,$$

y así sucesivamente.

La interpretación del momento de orden uno (media teórica de la variable) es muy clara, no siendo tan inmediata para los momentos de orden superior.

En la sección anterior demostrábamos que la esperanza de una variable aleatoria simétrica respecto del eje de ordenadas, es igual a cero. Obviamente, lo mismo ocurrirá para los momentos no centrales de orden impar. En efecto, si k es un número impar y X una variable continua, con función de densidad f_X,

$$\begin{aligned}\alpha_k = E(X^k) &= \int_{x=-\infty}^{\infty} x^k \cdot f_X(x)dx = \\ &= \int_{x=-\infty}^{0} x^k \cdot f_X(x)dx + \int_{x=0}^{\infty} x^k \cdot f_X(x)dx \,.\end{aligned}$$

Haciendo el cambio de variable $x = -t$ en la primera de las integrales anteriores,

$$\begin{aligned}\int_{x=-\infty}^{0} x^k \cdot f_X(x)dx &= \int_{t=\infty}^{0} (-t)^k \cdot f_X(-t)(-dt) = \\ &= -\int_{t=0}^{\infty} t^k \cdot f_X(t)dt \,.\end{aligned}$$

Por tanto, ambas integrales son iguales en valor absoluto, aunque de signo contrario y, consecuentemente, el momento α_k, si existe, será igual a cero.

Análogamente, se probaría el resultado para variables discretas.

Como ya se definió, *el momento centrado respecto de la media de orden k, con* $k = 0, 1, \ldots$, *o momento central de una variable aleatoria X, es el valor, si existe,*

$$\mu_k = E[(X - E(X))^k] = E[(X - \mu)^k] \,,$$

esto es, si X es discreta,

$$\mu_k = \sum_{i=1}^{\infty} (x_i - \mu)^k \cdot p[X = x_i]$$

y, si es continua,

$$\mu_k = \int_{-\infty}^{\infty} (x - \mu)^k \cdot f_X(x)dx \,.$$

Para $k = 0$ y $k = 1$, los momentos respecto de la media son siempre constantes:

$$\begin{aligned}\mu_0 &= E[(X - \mu)^0] = E(1) = 1 \\ \mu_1 &= E(X - \mu) = E(X) - \mu = 0 \,.\end{aligned}$$

Para $k = 2$, se obtiene el momento central más importante que recibe el nombre de varianza de la variable:

$$\mu_2 = E[(X - \mu)^2] \,.$$

A este momento, que también designaremos por $Var(X)$ o σ_x^2, le dedicaremos el siguiente subapartado.

Demostremos antes un resultado que relaciona los momentos centrales con los no centrales, y viceversa.

Proposición 6.2 *Sea X una variable aleatoria para la que existen los momentos, al menos, hasta un cierto orden k. Entonces,*

1. $\mu_k = \sum_{j=0}^{k} (-1)^j \cdot \binom{k}{j} \cdot \alpha_{k-j} \cdot \mu^j$

2. $\alpha_k = \sum_{j=0}^{k} \binom{k}{j} \cdot \mu_{k-j} \cdot \mu^j$

Demostración. Teniendo en cuenta el desarrollo del binomio $(X-\mu)^k$,

$$\mu_k = E[(X-\mu)^k] = E\left[\sum_{j=0}^{k}\binom{k}{j}\cdot X^{k-j}\cdot(-\mu)^j\right] = \sum_{j=0}^{k}(-1)^j\cdot\binom{k}{j}\cdot E(X^{k-j})\cdot\mu^j = \\ = \sum_{j=0}^{k}(-1)^j\cdot\binom{k}{j}\cdot\alpha_{k-j}\cdot\mu^j\ ,$$

con lo cual queda demostrado el primer resultado.

Utilizando un razonamiento análogo,

$$\begin{aligned}\alpha_k = E(X^k) = E[((X-\mu)+\mu)^k] &= E\left[\sum_{j=0}^{k}\binom{k}{j}\cdot(X-\mu)^{k-j}\cdot\mu^j\right] = \\ &= \sum_{j=0}^{k}\binom{k}{j}\cdot E[(X-\mu)^{k-j}]\cdot\mu^j = \\ &= \sum_{j=0}^{k}\binom{k}{j}\cdot\mu_{k-j}\cdot\mu^j\ ,\end{aligned}$$

demostrándose, así, el segundo resultado. ♣

Dado que los momentos no centrales son más fáciles de calcular que los momentos centrales, utilizaremos la primera expresión de la proposición anterior para obtener los segundos a partir de los primeros.

Veamos una propiedad de interés que cumplen ciertos momentos centrados respecto de la media para variables simétricas. En concreto, si X es simétrica respecto a un punto a cualquiera, los momentos centrales μ_k, con k impar, se anulan. En efecto, como ya probamos, si la función de probabilidad o de densidad de una variable aleatoria, X, es simétrica respecto de a, entonces, $E(X)$ vale, precisamente, a. Por tanto,

$$\mu_k = E[(X-\mu)^k] = E[(X-a)^k]\ .$$

Pero, si X es simétrica en torno al punto a, $Y = X - a$ es simétrica respecto del origen. En definitiva, para k impar,

$$\mu_k = E[(X-a)^k] = E(Y^k) = 0\ ,$$

como hemos visto en este mismo subapartado.

6.2.2 Varianza y coeficiente de variación

A continuación, estudiaremos el principal momento centrado respecto de la media, la varianza, así como otras medidas derivadas de ella.

Definición 6.5 *Sea X una variable aleatoria cuya esperanza es μ. Se llama varianza de X al número*

$$Var(X) = E[(X-\mu)^2] ,$$

que, si X es discreta, vale

$$Var(X) = \sum_{i=1}^{\infty}(x_i-\mu)^2 \cdot p[X=x_i]$$

y, si es continua,

$$Var(X) = \int_{-\infty}^{\infty}(x-\mu)^2 \cdot f_X(x)dx .$$

Como esperanza que es, la varianza de X, que designaremos indistintamente por $Var(X)$ o σ_X^2, presentará los mismos problemas de existencia que ya se han destacado en otras ocasiones.

Así como la esperanza es la media teórica de la distribución y, por tanto, bajo ciertas condiciones, un valor central representativo de la variable, la varianza es una medida absoluta de la dispersión de los valores de la variable respecto de su media. Su expresión no es otra que la de la media teórica de las diferencias cuadráticas respecto de la media. En este sentido, proporciona una medida global de las diferencias de los valores de la variable respecto de su esperanza, considerando éstas elevadas al cuadrado para evitar su signo, y, consecuentemente, que se contrarresten unas con otras (téngase en cuenta que $E(X-\mu)$ es siempre cero).

La finalidad de la varianza es doble: por un lado, valora el grado de dispersión de los valores de una distribución, permitiendo la comparación con otras distribuciones; por otro, da una medida de la representatividad de la esperanza como resumen global de la distribución.

Ejemplo 6.10. Un concesionario de automóviles vende 2 coches la mitad de los días y 16 la otra mitad. Otro concesionario vende 8 coches la mitad de los días y 10 la otra mitad. Estudiemos el número de coches que esperan vender cada uno de los concesionarios un día cualquiera. Sean X_1 y X_2 las variables que recogen el *número de coches vendidos* por cada uno de los establecimientos:

$$X_1: \left|\begin{array}{ll} 2 & p[X_1=2]=0.5 \\ 16 & p[X_1=16]=0.5 \end{array}\right. \qquad X_2: \left|\begin{array}{ll} 8 & p[X_2=8]=0.5 \\ 10 & p[X_2=10]=0.5 \end{array}\right.$$

Entonces,

$$E(X_1) = 2\cdot 0.5 + 16\cdot 0.5 = 9 \text{ coches}$$

y, análogamente,

$$E(X_2) = 9 \text{ coches.}$$

Ambos concesionarios esperan vender nueve coches diarios aunque, a la vista de los valores de las variables, parece lógico pensar que esa cantidad no es igualmente representativa en ambos casos. El cálculo de las varianzas confirma esta idea:

$$Var(X_1) = (2-9)^2 \cdot 0.5 + (16-9)^2 \cdot 0.5 = 49$$

y

$$Var(X_2) = (8-9)^2 \cdot 0.5 + (10-9)^2 \cdot 0.5 = 1\ .$$

La dispersión de los valores de X_1 es mucho mayor que la de X_2; por tanto, el número esperado de coches vendidos es mucho más representativo de la distribución de las ventas para el segundo concesionario que para el primero. ◁

La varianza es siempre un valor no negativo, pues se trata de la esperanza del cuadrado de $(X-\mu)$, es decir, de la esperanza de una variable no negativa. En muchas ocasiones, consideraremos la raíz cuadrada positiva de la varianza: la **desviación típica** o **desviación estándar**, σ_X. Esta medida absoluta de dispersión presenta la ventaja de estar dada en las mismas unidades de la variable, frente a la varianza que viene medida en dichas unidades al cuadrado.

Seguidamente, demostraremos algunas de las principales propiedades de la varianza, suponiendo que ésta existe.

P1. Sea X una variable aleatoria. Entonces,

$$Var(X) = E(X^2) - [E(X)]^2\ .$$

Demostración. Desarrollando el binomio $(X-\mu)^2$,

$$Var(X) = E[(X-\mu)^2] = E[X^2 - 2\mu \cdot X + \mu^2] = E(X^2) - \mu^2\ . \clubsuit$$

Dicha propiedad es un caso particular de la proposición enunciada y demostrada anteriormente sobre la relación entre momentos centrales y no centrales. En los cálculos de la varianza será más práctico utilizar esta expresión que la que resulta de su definición.

P2. La varianza de una variable es cero si, y sólo si, esta variable es degenerada, es decir, es constante.

Demostración. Si $Var(X) = 0$, todos los sumandos $(x_i - \mu)^2 \cdot p[X = x_i]$ se anularán. Por tanto, x_i será igual a μ para todos los valores de i. Entonces, X será una variable que toma siempre el mismo valor, μ, es decir, X será una variable degenerada.

Evidentemente, no tiene sentido plantearse la demostración anterior para variables continuas (sería absurda dicha suposición para llegar a la conclusión de que es una variable degenerada y, por tanto, discreta).

Por otra parte, si X es degenerada, toma siempre el mismo valor K, y, consecuentemente, $p[X = K] = 1$. Ya hemos visto en el apartado anterior que, en este caso, $E(X) = K$. Calculemos su varianza:

$$Var(X) = \sum_{i=1}^{\infty}(x_i - \mu)^2 \cdot p[X = x_i] = (K - K)^2 \cdot p[X = K] = 0 \,. \clubsuit$$

Naturalmente, el interés de la Estadística no está en las variables degeneradas. Ahora bien, la propiedad demostrada nos indica que en la medida en que la varianza de una variable se acerca (se aleja) del cero, esta variable presentará menos (más) variabilidad.

P3. El momento centrado respecto de una constante C de orden dos se hace mínimo cuando C toma el valor μ. Es decir, para cualquier valor C,

$$E[(X - C)^2] \geq Var(X) \,,$$

alcanzándose la igualdad cuando $C = \mu$.

Demostración.

$$\begin{aligned} E[(X - C)^2] &= E[(X - \mu) + (\mu - C)]^2 = \\ &= E[(X - \mu)^2] + (\mu - C)^2 - 2(\mu - C)E(X - \mu) = \\ &= VarX + (\mu - C)^2 \,. \end{aligned}$$

Es evidente que $E[(X - C)^2]$ se hace mínima cuando $C = \mu$, es decir, cuando coincide con la varianza. $\clubsuit$

El resultado anterior que, como la mayoría de los resultados referentes a las medidas teóricas, tiene su paralelo en las medidas empíricas, recibe el nombre de teorema de KÖNIG.

P4. Sea X una variable aleatoria y b una constante cualquiera. Entonces,

$$Var(X + b) = Var(X)$$

y, por tanto,

$$\sigma_{X+b} = \sigma_X \,.$$

Demostración. Sin más que tener en cuenta la definición de varianza de una variable,

$$Var(X + b) = E[((X + b) - E(X + b))^2] = E[(X - E(X))^2] = Var(X) \,. \clubsuit$$

P5. Sea X una variable aleatoria y a una constante cualquiera. Entonces,

$$Var(a \cdot X) = a^2 \cdot Var(X)$$

y, por tanto,

$$\sigma_{a \cdot X} = |a| \cdot \sigma_X \,.$$

Demostración. Razonando como en la propiedad anterior,

$$Var(a \cdot X) = E[(a \cdot X - E(a \cdot X))^2] = E[a^2(X - E(X))^2] = a^2 \cdot Var(X) \,. \clubsuit$$

Como consecuencia inmediata de este resultado, tomando $a = -1$, se deduce que

$$Var(-X) = Var(X) \,.$$

Resumiendo las dos propiedades anteriores, obtenemos

$$Var(a \cdot X + b) = a^2 \cdot Var(X)$$

y, por tanto,

$$\sigma_{a \cdot X + b} = |a| \cdot \sigma_X \,.$$

Como ya comentamos, el cambio de origen y de escala es un caso particular de la transformación anterior. En este caso, la varianza de la variable transformada,

$$Y = \frac{X - o}{e} \,,$$

será, pues,

$$Var(Y) = \frac{Var(X)}{e^2} \,,$$

es decir,

$$\sigma_Y^2 = \frac{\sigma_X^2}{e^2} \,.$$

La desviación típica de Y valdrá, consecuentemente,

$$\sigma_Y = \frac{\sigma_X}{e} \,.$$

Por tanto, un cambio en el origen de la variable no influye sobre la varianza, dado que este cambio mueve de posición la variable, pero no dispersa o concentra sus valores. Sin embargo, la varianza no es invariante frente a cambios de escala. Al dividir todos los valores de una variable por un número, e, éstos se concentran (si dicho número es mayor que la unidad) o se dispersan (si es menor que la unidad).

La varianza (o la desviación típica) es una medida absoluta de la dispersión de la variable y, como tal, está influida por el tamaño de los valores de la misma. Para evitar esta influencia, debemos valorar esta medida de dispersión respecto del tamaño de la variable (medido en términos del valor esperado, μ_X). Ello nos lleva a definir, para una variable de media μ_X y varianza σ_X^2, una medida relativa de dispersión, el **coeficiente de variación**:

$$CV_X = \frac{\sigma_X}{\mu_X} \,.$$

Teniendo en cuenta que el tamaño de la variable está medido en términos de su esperanza, μ_X, el coeficiente de variación sólo tendrá sentido cuando μ_X sea

una medida razonable del tamaño de X. Como es obvio, esto ocurrirá, únicamente, cuando X sea una variable positiva. Si la variable tomase valores positivos y negativos, la media quedaría compensada, más o menos, por dichos valores, no proporcionando, por tanto, una idea del tamaño de X. Una situación límite la encontraremos cuando tengamos una variable que tome valores positivos y negativos, siendo su media cero: lógicamente, $\mu_X = 0$ no sirve, en absoluto, para evaluar el tamaño de la variable; por otra parte, el coeficiente de variación no estaría, ni siquiera, definido en este caso.

Estudiemos el comportamiento del coeficiente de variación de una variable aleatoria frente a los cambios de origen y de escala. Si X es una variable aleatoria, e Y su transformada,

$$Y = \frac{X - o}{e} ;$$

entonces, la esperanza de Y vale

$$\mu_Y = \frac{\mu_X - o}{e}$$

y su desviación típica,

$$\sigma_Y = \frac{\sigma_X}{e} .$$

En consecuencia, si μ_X y μ_Y son no nulos,

$$CV_Y = \frac{\sigma_Y}{\mu_Y} = \frac{\frac{\sigma_X}{e}}{\left(\frac{\mu_X - o}{e}\right)} = \frac{\sigma_X}{\mu_X - o} .$$

Por tanto, $CV_Y \neq CV_X$, salvo cuando o es cero.

Este resultado indica que el coeficiente de variación no se ve afectado por los cambios de escala, pero, por otro lado, dicho resultado muestra que el coeficiente de variación se modifica arbitrariamente si se realiza un cambio de origen. Esto pudiera parecer poco razonable, ya que el traslado de una variable aleatoria no debiera alterar su dispersión, pues no se modifican ni las posiciones relativas de los distintos valores, ni sus probabilidades. Ahora bien, adviértase que el coeficiente de variación es una medida de dispersión relativa, pero esta relatividad está referida al tamaño de la variable (medido en términos de su esperanza), y resulta lógico que el coeficiente de variación cambie cuando se modifique el tamaño de la variable al hacer un cambio de origen.

Por ejemplo, si X es una variable aleatoria discreta que toma sólo los valores 0 y 1 con la misma probabilidad, su coeficiente de variación será:

$$CV_X = \frac{\sigma_X}{\mu_X} = \frac{1/2}{1/2} = 1 .$$

Si trasladamos diez unidades la variable, esto es, si consideramos $Y = X + 10$, la nueva variable tomará los valores 10 y 11, con idéntica probabilidad. Entonces, su coeficiente de variación será:

$$CV_Y = \frac{\sigma_Y}{\mu_Y} = \frac{\frac{1}{2}}{\frac{1}{2} + 10} = \frac{1}{21} .$$

La medida de dispersión relativa se ha divido por 21: las distancias entre los valores de la variable no han cambiado, pero sí el tamaño de la misma, antes entre 0 y 1, y, ahora, entre 10 y 11. Obsérvese que, en una *situación límite*, si trasladamos la variable 0.5 unidades hacia la izquierda, es decir, si definimos $Z = X - 0.5$, entonces, $\mu_Z = 0$ y, por tanto, el coeficiente de variación no estaría definido.

En consecuencia, el coeficiente de variación se emplea sólo para comparar dispersiones de variables aleatorias positivas, siempre que no sean objeto de traslación. Se puede modificar arbitrariamente su escala de medida (por ejemplo, una variable aleatoria monetaria se podría medir en pesetas, millones o céntimos), pero el origen debe conservarse inalterado.

Excepcionalmente, se emplea también esta medida para variables aleatorias negativas. En este caso, se considera definido el coeficiente de variación en valor absoluto, pues, ahora, $|\mu_X|$ sí puede ser considerado una buena medida del tamaño de la variable. Así,

$$CV_X = \left|\frac{\sigma_X}{\mu_X}\right| = \frac{\sigma_X}{|\mu_X|} .$$

En muchas ocasiones, el coeficiente de variación que, obviamente, es una medida adimensional, se utiliza en porcentaje.

Ejemplo 6.11. Los obreros de una empresa se acogen a tres categorías salariales: un tercio de ellos cobra 50 mil pesetas al mes, otro tercio 100 mil y, por último, el resto 150 mil. Por su parte, los directivos de dicha empresa también se agrupan en tres categorías salariales de igual tamaño, con salarios de 450, 500 y 550 mil pesetas. Calculemos y comparemos el salario esperado y su varianza, para los obreros, X_1, y para los directivos, X_2.

X_1 :		
	50	$p[X_1 = 50] = 1/3$
	100	$p[X_1 = 100] = 1/3$
	150	$p[X_1 = 150] = 1/3$

X_2 :		
	450	$p[X_2 = 450] = 1/3$
	500	$p[X_2 = 500] = 1/3$
	550	$p[X_2 = 550] = 1/3$

Resulta inmediato que

$$E(X_1) = 100 \text{ mil pesetas}, \quad Var(X_1) = 1666.67$$
$$E(X_2) = 500 \text{ mil pesetas} \quad \text{y} \quad Var(X_2) = 1666.67 .$$

Ambas variables tienen la misma varianza, dado que en ambos casos los valores salariales de las dos terceras partes del colectivo difieren en 50 mil pesetas del salario esperado (100 mil pesetas en un caso y 500 mil en el otro), mientras la tercera parte restante tiene un salario que coincide con la media. Ahora bien, es evidente que esas 50 mil pesetas de desviación respecto de las medias, no tienen el mismo significado cuando hablamos de valores alrededor de 100 mil pesetas que cuando hablamos de valores alrededor de 500 mil. En el caso de los obreros, una categoría (sueldos medios) tiene salarios superiores en un 100 % a los de la categoría inferior (sueldos bajos), mientras que, para la categoría superior (sueldos altos), ese incremento es del 200 %. En cambio, entre los directivos, los salarios de cada categoría se incrementan en poco más del 11 % respecto a los de la categoría anterior. Esta dispersión relativa de los salarios, mucho mayor entre los obreros que entre los directivos, queda reflejada en el coeficiente de variación:

$$CV_{X_1} = \frac{40.82}{100} = 0.408 \quad \text{y} \quad CV_{X_2} = \frac{40.82}{500} = 0.082 ,$$

es decir, entre los obreros la desviación típica representa el 40.8 % del salario esperado, mientras que, para los directivos, este porcentaje es mucho más pequeño, en concreto del 8.2 %. ◁

6.2.3 Tipificación de una variable

En muchas ocasiones, se mide una misma magnitud con distintas unidades o en distintas situaciones (test de inteligencia, productividad, etc.). De cara a poder comparar los resultados obtenidos, es conveniente, a veces, homogeneizarlos. Dicho proceso recibe el nombre de tipificación.

Dada una variable X, de media μ_X y varianza σ_X^2, decimos que Z es la variable X tipificada, si

$$Z = \frac{X - \mu_X}{\sigma_X} .$$

Por tanto, tipificar una variable consiste en transformarla mediante un cambio de origen y de escala en el que $o = \mu_X$ y $e = \sigma_X$. La variable transformada, Z, queda *despojada* de los efectos de las unidades de medida; entonces, Z podrá ser comparada con otras variables también tipificadas.

¿Cuáles serán la esperanza y la varianza de la nueva variable? Teniendo en cuenta cómo repercuten los cambios de origen y de escala en la esperanza y la varianza, tendremos que

$$\mu_Z = \frac{\mu_X - \mu_X}{\sigma_X} = 0$$

y

$$\sigma_Z^2 = \frac{\sigma_X^2}{\sigma_X^2} = 1 .$$

Por tanto, la variable tipificada tiene media 0 y varianza 1.

Este proceso de tipificación es de gran importancia en diversos desarrollos estadísticos que impliquen tratamientos de múltiples variables aleatorias.

Ejemplo 6.12. Las puntuaciones en estadística de los alumnos de una clase, X_1, se ajustan a la siguiente función de densidad:

$$f_{X_1}(x) = \begin{cases} x/25 & 0 \le x < 5 \\ (10 - x)/25 & 5 \le x \le 10 \\ 0 & \text{en el resto} \end{cases}$$

Tipifiquemos esta variable. Para ello, calculemos la esperanza y la varianza de X_1. En primer lugar,

$$E(X_1) = \int_0^{10} x \cdot f_{X_1}(x)dx = \int_0^5 x \cdot \frac{x}{25}dx + \int_5^{10} x \cdot \left(\frac{10 - x}{25}\right) dx = 5 \text{ puntos.}$$

Para hallar la varianza, obtengamos, previamente, el momento respecto del origen de orden 2:

$$E(X_1^2) = \int_0^{10} x^2 \cdot f_{X_1}(x)dx = \int_0^5 x^2 \cdot \frac{x}{25}dx + \int_5^{10} x^2 \cdot \left(\frac{10 - x}{25}\right) dx = \frac{75}{2} ,$$

y, por tanto,

$$Var(X_1) = \frac{75}{2} - 5^2 = \frac{25}{2} .$$

La variable tipificada es

$$Z_1 = \frac{X_1 - 5}{\sqrt{25/2}} = \frac{\sqrt{2}(X_1 - 5)}{5} .$$

Mediante un sencillo cambio de variable, el lector puede calcular la función de densidad de la variable tipificada.

Por otra parte, la puntuación en matemáticas, X_2, sigue una distribución de media 4 y varianza 8. La variable X_2 tipificada será:

$$Z_2 = \frac{X_2 - 4}{2\sqrt{2}} .$$

Un alumno obtuvo un 7 en estadística y un 6 en matemáticas. ¿Quiere esto decir que dicho alumno, comparado con el resto de sus compañeros, está mejor preparado en estadística que en matemáticas? En principio, no se pueden comparar, sin más, las dos puntuaciones, pues provienen de variables con distinta media y varianza. Si se quieren comparar, habrá que tipificarlas previamente. Así, el 7 en estadística se transforma en

$$\frac{\sqrt{2}(7-5)}{5} = 0.57 ,$$

mientras que el 6 en matemáticas, en

$$\frac{6-4}{2\sqrt{2}} = 0.71 .$$

De las puntuaciones tipificadas deducimos que el alumno, comparado con el resto de los compañeros de clase, está mejor preparado en matemáticas que en estadística. En otros términos, que el 6 en matemáticas es *más* que el 7 en estadística. ◁

6.3 Otras medidas de posición, dispersión y forma

La esperanza y la varianza son medidas teóricas de posición (en concreto, de centralización) y dispersión, respectivamente. Pero, aunque son las más generalizadas, no son las únicas medidas que se pueden utilizar.

Las medidas de posición sitúan la distribución de probabilidad en la recta real; dentro de éstas, las medidas de posición central proporcionan un valor central de la variable. Por su parte, las medidas de dispersión indican el grado de concentración de los valores de la variable alrededor de una medida de posición central, dando, a su vez, una idea de la representatividad de esta medida de centralización como resumen global de la variable.

Entre otras medidas de posición, las más utilizadas corresponden a la familia de los cuantiles.

Definición 6.6 *Sea X una variable aleatoria. Un cuantil de orden q $(0 \leq q \leq 1)$ es cualquier valor, x_q, tal que*

$$p[X \leq x_q] \geq q \quad y \quad p[X \geq x_q] \geq 1 - q .$$

Los cuantiles dividen a la variable en dos partes; a la izquierda, el conjunto de valores inferiores o iguales a x_q cuya probabilidad es mayor o igual que q, y, a la derecha, el conjunto de valores superiores o iguales a x_q con una probabilidad de al menos $1-q$. De la definición anterior, el lector puede deducir que el cuantil de orden q no tiene por qué ser único.

Ejemplo 6.13. Sea X una variable aleatoria discreta cuya función de probabilidad se describe a continuación:

$$X: \left| \begin{array}{ll} 0 & p[X=0]=0.1 \\ 1 & p[X=1]=0.3 \\ 2 & p[X=2]=0.2 \\ 3 & p[X=3]=0.4 \end{array} \right.$$

Calculemos los cuantiles de orden 0.4 y 0.5. El primero de ellos, $x_{0.4}$, será el 1, el 2, y cualquier valor intermedio, puesto que

$$p[X \leq 1] = 0.4 \geq 0.4\,, \qquad p[X \geq 1] = 0.9 \geq 1 - 0.4$$

y

$$p[X \leq 2] = 0.6 \geq 0.4\,, \qquad p[X \geq 2] = 0.6 \geq 1 - 0.4\,.$$

El segundo cuantil, $x_{0.5}$, será el 2, pues

$$p[X \leq 2] = 0.6 \geq 0.5\,, \qquad p[X \geq 2] = 0.6 \geq 1 - 0.5\,,$$

es decir, el 2 es el único valor de la variable que deja, tanto a su derecha como a su izquierda, al menos el 50 % de la probabilidad. ◁

En el caso en que las variables sean continuas, la definición anterior se simplifica, dado que el aporte de probabilidad es continuo, y no a saltos, como en las variables discretas.

Definición 6.7 *Sea X una variable aleatoria continua. Un cuantil de orden q es cualquier valor, x_q, tal que*

$$p[X \leq x_q] = q\,.$$

Es decir, el cuantil de orden q deja a la izquierda una cantidad q de probabilidad. Por tanto, el cuantil es un valor para el cual la función de distribución vale q, esto es,

$$F_X(x_q) = q\,.$$

En este caso, el cuantil x_q será único salvo que la función de distribución sea constante, e igual a q, en un intervalo; entonces, x_q será cualquier valor de ese intervalo.

Ejemplo 6.14. Sea X una variable aleatoria continua cuya función de distribución es la siguiente:

$$F_X(x) = \begin{cases} 1 - e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Hallemos los cuantiles $x_{0.25}, x_{0.50}$ y $x_{0.75}$.

De la expresión de la función de distribución deducimos que

$$F_X(x_{0.25}) = 1 - e^{-x_{0.25}} = 0.25 .$$

Por tanto,

$$e^{-x_{0.25}} = 0.75 .$$

Así,

$$x_{0.25} = -\ln(0.75) = 0.29 .$$

De igual manera,

$$x_{0.50} = -\ln(0.5) = 0.69$$

y

$$x_{0.75} = -\ln(0.25) = 1.39 .$$

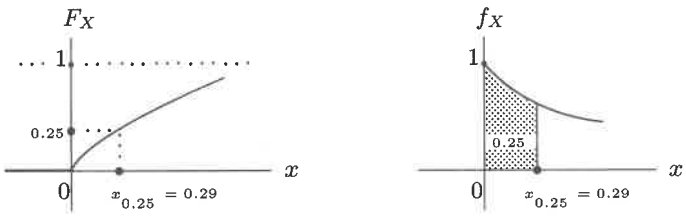


FIG. 6.6.

Los gráficos anteriores representan el significado de uno de los tres cuantiles, el cuantil $x_{0.25}$, en las funciones de distribución y de densidad. ◁

Dentro de los cuantiles, interesan fundamentalmente aquellos que dividen el total de probabilidad en mitades, cuartas, décimas o centésimas partes. Así, obtendríamos:

- Cuantil de orden 0.5: **mediana** (Me).
- Cuantiles de orden 0.25 y 0.75: primer y tercer **cuartil** (C_1 y C_3).
- Cuantiles de orden $0.1, 0.2, \ldots$ y 0.9: primer, segundo, ... y noveno **decil**, denotados como $d_1, d_2, \ldots$ y d_9.
- Cuantiles de orden $0.01, 0.02, \ldots$ y 0.99: primer, segundo, ... y nonagésimo nono **percentil** ($p_1, p_2, \ldots$ y p_{99}).

Los cuantiles sitúan las distintas fracciones de la variable en la recta real. En este sentido, pueden interpretarse, bien en solitario, bien conjuntamente. Así, en las distribuciones de las rentas familiares de un cierto país, es usual proporcionar simultáneamente todos los deciles. Cada par de deciles consecutivos proporciona un intervalo de renta (decila) que acoge un 10 % de las familias. Así, la primera decila, $(0, d_1)$, incluye el 10 % de las familias con rentas más bajas; la segunda decila, (d_1, d_2), incluye al 10 % con rentas inmediatamente superiores, y así sucesivamente.

Todo ello proporciona una buena idea de la distribución de la variable, en este caso, de la renta familiar.

El cuantil más relevante es la mediana, que divide a la variable en dos partes de probabilidad igual (con las consideraciones pertinentes si la variable es discreta). Este cuantil de orden 0.5 constituye una buena medida de posición central, que en ciertas situaciones es preferible a la media.

Ejemplo 6.15. Consideremos, de nuevo, la variable *número de hijos por familia* del ejemplo 6.5.

X_i	$p[X = x_i]$
0	0.4
1	0.3
2	0.2
23	0.1

Como ya vimos, el número esperado de hijos por familia es 3, valor que en absoluto representa a la distribución. La mediana de X es 1, ya que 1 es el valor de la variable que deja a su izquierda, y a su derecha, al menos la mitad de la probabilidad. En concreto,

$$p[X \leq 1] = 0.7 \geq 0.5 \quad \text{y} \quad p[X \geq 1] = 0.6 \geq 0.5 \,.$$

Sin más que observar la distribución de X, deducimos que la mediana, 1 hijo, es un valor mucho mas representativo de la misma que la media, 3 hijos. ◁

Otra medida de posición central de menor interés es la **moda** (Mo). Se define como el valor de la variable en el que la función de probabilidad o la función de densidad alcanza el máximo absoluto. Intuitivamente, es el valor más probable, pero no olvidemos que, para las variables continuas, todos los valores tienen una misma probabilidad: cero.

Para las variables discretas, el cálculo de la moda (cuando no surja de la mera observación de la función de probabilidad), se realizará buscando previamente los máximos relativos de dicha función; es decir, hallando los valores x_i que verifiquen las dos desigualdades siguientes:

$$\begin{aligned} p[X = x_i] &\geq p[X = x_{i-1}] \\ p[X = x_i] &\geq p[X = x_{i+1}] \,. \end{aligned}$$

Para las variables continuas, el problema se traduce en encontrar los valores que anulen la primera derivada de la función de densidad, y hagan negativa la segunda, eligiendo, entre ellos, aquel en el que $f_X(x)$ alcance el mayor valor.

Mientras que la esperanza y la mediana (salvo en ciertas circunstancias esta última) toman un único valor, la moda puede ser múltiple, perdiendo, en alguno de estos casos, su sentido como valor central.

En cuanto a otras medidas de dispersión a añadir a las ya vistas, cabe citar las desviaciones absolutas medias respecto de alguna medida de centralización. Entre éstas, la más utilizada es

$$d_{Me} = E[|X - Me|] \,,$$

denominada **desviación absoluta media respecto a la mediana**.

Por otro lado, a partir de los cuantiles, podemos construir diversas medidas de dispersión. En concreto, el **recorrido intercuartílico**:

$$R_c = x_{0.75} - x_{0.25} = C_3 - C_1 \ ,$$

medida que indica la longitud del intervalo intercuartílico, en el que se encuentra el 50 % central de los valores de la variable, desechando los valores más o menos *raros* (el 25 % de los más pequeños y el 25 % de los más grandes).

Si queremos considerar el 100 % de los valores, y no solamente la mitad, tendremos el **recorrido** de la variable:

$$R = \max(X) - \min(X) \ ,$$

pudiendo considerarse el máximo y el mínimo valor de la variable como cuantiles de orden 1 y 0, respectivamente. Esta medida es menos precisa que la anterior, al recoger valores extraños que pueden distorsionar la verdadera dispersión de la variable. Obsérvese que, tanto el recorrido como el recorrido intercuartílico, miden la dispersión en sí misma, y no respecto a una medida de centralización, como ocurre en la varianza o en la desviación absoluta media respecto de la mediana.

Además de las medidas de posición y dispersión, existen otras medidas que informan sobre el perfil de la función de probabilidad o de densidad de la variable; en concreto, sobre el grado de simetría y apuntamiento o aplastamiento de dichas funciones. La finalidad de estas medidas, **medidas de forma**, está en conocer si las probabilidades se distribuyen simétricamente alrededor de la media y si dicha distribución tiene un comportamiento similar a la distribución de probabilidad más importante de la Estadística: la distribución normal. A continuación, veremos dos de estas medidas debidas a R. A. FISHER.

Como ya hemos visto, las variables cuya función de probabilidad o de densidad es simétrica respecto de un punto, tienen su esperanza en dicho punto. En este caso, las diferencias entre cada valor y la media elevadas a potencias impares se contrarrestan, resultando todos sus momentos centrales de orden impar nulos (el de orden uno siempre lo es).

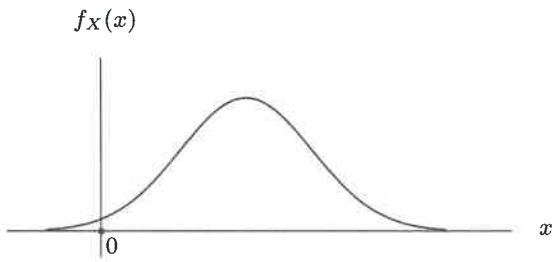


FIG. 6.7.

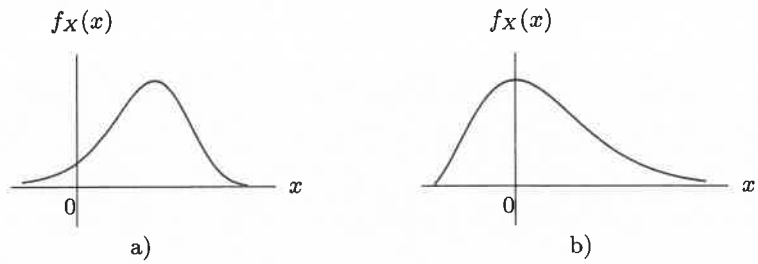


FIG. 6.8.

Basándose en lo anterior, FISHER propone el siguiente **coeficiente de asimetría**:

$$\gamma_1 = \frac{\mu_3}{\sigma^3} .$$

Dicho coeficiente, adimensional e invariante ante cambios de origen y escala, vale cero cuando la variable es simétrica (figura 6.7).

Para distribuciones unimodales con desviación a la derecha (figura 6.8.a), el coeficiente γ_1 toma valores positivos, pues las desviaciones de los valores de la derecha (diferencias positivas) tienen más peso en μ_3 que las de los valores de la izquierda (diferencias negativas). Para distribuciones unimodales con desviación a la izquierda (figura 6.8.b) ocurre lo contrario, tomando γ_1 valores negativos[1].

Trataremos, ahora, de estudiar el aplastamiento o apuntamiento de la función de densidad de una variable aleatoria continua, comparándola con el perfil de una distribución patrón con la misma media y varianza. Dicha distribución es la distribución normal de media μ y varianza σ^2, que estudiaremos en profundidad en el capítulo 10. Su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty .$$

El perfil de esta función se muestra en la figura 6.9.

Para esta distribución, el momento de orden cuatro respecto de la media vale:

$$\mu_4 = 3 \cdot \sigma^4 .$$

Cuando las distribuciones son unimodales y más o menos simétricas, comparamos el perfil de su función de densidad con el de la normal de la misma media y varianza, viendo si aquél es más apuntado (más probabilidad en el centro de la distribución) o más aplastado (más probabilidad en las colas) que el de la normal

1. Nótese que, aunque la simetría o asimetría de la distribución determina el valor del coeficiente γ_1, el hecho de que, por ejemplo, γ_1 tome el valor cero, no significa que la distribución sea siempre simétrica, si bien es cierto que lo será en la mayoría de las situaciones de interés. Esto sugiere una interpretación cautelosa de los valores del coeficiente γ_1.

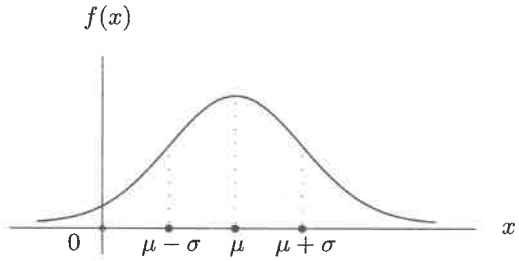


FIG. 6.9.

correspondiente. En virtud de lo anterior, FISHER propone el siguiente **coeficiente de curtosis** o **apuntamiento**:

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \, .$$

Este coeficiente, adimensional e invariante ante cambios de origen y de escala, toma el valor cero cuando la distribución presenta el mismo perfil que la normal correspondiente (distribuciones mesocúrticas).

Si la distribución es más apuntada que la normal con la misma media y varianza, entonces γ_2 es positivo (distribución leptocúrtica); en el caso de ser más aplastada, γ_2 es negativo (distribución platicúrtica)[2] (véase la figura 6.10).

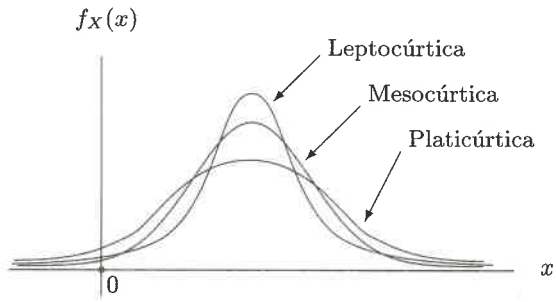


FIG. 6.10.

Adviértase que la distribución que se toma como patrón, la normal, es una distribución continua, por tanto, con función de densidad. Si estudiamos el apuntamiento de una distribución continua, la comparación con la normal no presentará problemas. Ahora bien, si la variable es discreta, no tiene sentido comparar el perfil de su función de probabilidad con el de la función de densidad normal. Aun así, el

2. Las mismas reservas efectuadas para el coeficiente γ_1 en la nota a pie de página anterior son válidas para γ_2.

estudio del aplastamiento o apuntamiento se puede realizar, siempre y cuando se interprete el coeficiente, no en términos gráficos, sino en el sentido del reparto de la probabilidad en la recta real. Así, si γ_2 es cero, la variable discreta *dispondrá* su probabilidad en la recta real de forma similar a la normal. En cambio, si γ_2 es mayor que cero, la variable acumulará más probabilidad en el centro y menos en las colas que la normal correspondiente; de forma análoga, interpretaremos valores negativos de γ_2.

Ejemplo 6.16. Sea una variable aleatoria continua, X, cuya función de densidad es:

$$f_X(x) = \begin{cases} 12x^2(1-x) & 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Vamos a obtener sus principales características numéricas. Así, la esperanza será:

$$E(X) = \int_{x=0}^{1} x \cdot 12x^2(1-x)\, dx = \frac{12}{20} = 0.6\ .$$

Para calcular la mediana, hallemos previamente la función de distribución. Para valores de x entre cero y uno,

$$F_X(x) = \int_{t=0}^{x} 12t^2(1-t)\, dt = 4x^3 - 3x^4\ .$$

Por tanto,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 4x^3 - 3x^4 & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

La mediana verificará:

$$F_X(Me) = 0.5\ ,$$

es decir, la mediana será el valor que resuelve la ecuación:

$$4x^3 - 3x^4 - 0.5 = 0\ .$$

Dicho valor es 0.614.

Por otro lado, la moda será el máximo absoluto de la función de densidad. Derivando esta función e igualando a cero, obtenemos:

$$f'_X(x) = 12(2x - 3x^2) = 12x(2 - 3x) = 0\ ,$$

siendo $x = 0$ y $x = \frac{2}{3}$ los valores que anulan la derivada. Ahora bien, $f''_X\left(\frac{2}{3}\right) = -24$. Por consiguiente, la función de densidad alcanza el máximo en $x = \frac{2}{3}$, siendo este valor la moda de la distribución.

Obsérvese que las tres medidas de centralización no coinciden (figura 6.11); en concreto,

$$E(X) < Me < Mo\ .$$

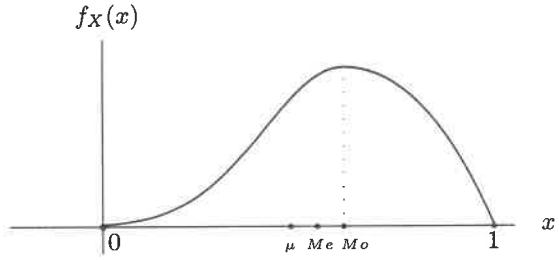


FIG. 6.11.

Obtengamos la varianza de la distribución. Previamente,

$$E(X^2) = \int_0^1 x^2 \cdot 12x^2(1-x)\,dx = 0.4\,.$$

De aquí,

$$Var(X) = E(X^2) - (E(X))^2 = 0.04\,,$$

siendo la desviación típica 0.2.

El coeficiente de variación será

$$CV_X = \frac{\sigma_X}{\mu_X} = \frac{1}{3}\,,$$

es decir, la desviación es el 33.3 % de la esperanza.

Calculemos la desviación absoluta media respecto a la mediana:

$$d_{Me} = E[|X - Me|] = E[|X - 0.614|] = \int_0^1 |x - 0.614| \cdot f_X(x)dx\,.$$

Descomponiendo la integral obtenemos:

$$d_{Me} = \int_0^{0.614} -(x - 0.614) \cdot 12x^2(1-x)dx + \int_{0.614}^1 (x - 0.614) \cdot 12x^2(1-x)dx = 0.17\,.$$

El cálculo de d_{Me}, es decir, de la esperanza de $Y = |X - Me|$ también podría realizarse mediante la obtención previa de la función de densidad de Y.

Para obtener el recorrido intercuartílico, R_c, hay que hallar los dos cuartiles C_1 y C_3 que verifican

$$F_X(C_1) = 0.25$$

y

$$F_X(C_3) = 0.75\,,$$

o, lo que es lo mismo, hay que resolver las dos ecuaciones siguientes:

$$4C_1^3 - 3C_1^4 = 0.25$$

y

$$4C_3^3 - 3C_3^4 = 0.75\,.$$

La resolución de estas ecuaciones lleva a los valores $C_1 = 0.456$ y $C_3 = 0.757$, siendo, por tanto, $R_c = 0.301$. Es decir, en un intervalo de longitud 0.301 se encuentra el 50 % central de la variable.

Estudiemos la forma de la distribución, esto es, la simetría y el apuntamiento de su función de densidad. En las expresiones de los respectivos coeficientes, aparecen los momentos centrales μ_3 y μ_4, que hallaremos a partir de los momentos respecto del origen.

$$\alpha_k = E(X^k) = \int_0^1 x^k \cdot 12x^2(1-x)\,dx = \frac{12}{(k+3)(k+4)} .$$

Es decir,

$$\alpha_1 = \frac{3}{5} , \quad \alpha_2 = \frac{2}{5} , \quad \alpha_3 = \frac{2}{7} , \quad \alpha_4 = \frac{3}{14} , \cdots$$

El primer resultado de la proposición 6.2 permite obtener los momentos centrales como combinación lineal de los no centrales:

$$\begin{aligned} \mu_2 &= \sigma^2 = \alpha_2 - \alpha_1^2 = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25} , \\ \mu_3 &= \alpha_3 - 3\alpha_2 \cdot \alpha_1 + 3\alpha_1 \cdot \alpha_1^2 - \alpha_1^3 = -\frac{2}{875} \qquad \text{y} \\ \mu_4 &= \alpha_4 - 4\alpha_3 \cdot \alpha_1 + 6\alpha_2 \cdot \alpha_1^2 - 4\alpha_1 \cdot \alpha_1^3 + \alpha_1^4 = \frac{33}{8750} . \end{aligned}$$

El coeficiente de asimetría es

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = -0.29 ,$$

valor que confirma una débil desviación a la izquierda, como se refleja en la figura 6.11.

Del mismo modo, el coeficiente de curtosis es

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = -0.64 ,$$

que pone de manifiesto un grado de apuntamiento menor que el de la distribución normal correspondiente, es decir, la distribución es platicúrtica. ◁

6.4 Teorema de Markov y desigualdad de Chebyshev

Los dos teoremas siguientes permiten calcular cotas superiores de ciertas probabilidades, cuando conocemos, únicamente, la esperanza y la varianza de la distribución. El primero de ellos recibe el nombre de **teorema de Markov**.

Teorema 6.1 *Sea X una variable aleatoria no negativa cuya esperanza existe, y sea ε un número positivo. Entonces,*

$$p[X \geq \varepsilon] \leq \frac{E(X)}{\varepsilon} .$$

Demostración. Si X es una variable continua,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^{\varepsilon} x \cdot f(x)dx + \int_{\varepsilon}^{\infty} x \cdot f(x)dx = \\ &= \int_{\varepsilon}^{\infty} x \cdot f(x)dx \geq \int_{\varepsilon}^{\infty} \varepsilon \cdot f(x)dx = \varepsilon \cdot p[X \geq \varepsilon] , \end{aligned}$$

de donde obtenemos el resultado.

De igual manera, demostraríamos el teorema para variables discretas. ♣

Una consecuencia inmediata del teorema anterior es la llamada **desigualdad de Chebyshev**, que proporciona una cota superior de la probabilidad del conjunto de valores de la variable que difieren de la media en, al menos, una cierta cantidad.

Teorema 6.2 *Sea X una variable aleatoria de media μ y varianza σ^2, y sea K un número positivo. Entonces,*

$$p[|X-\mu| \geq K] \leq \frac{\sigma^2}{K^2} .$$

Demostración. Aplicamos el teorema de MARKOV a la variable no negativa $(X-\mu)^2$:

$$p[(X-\mu)^2 \geq \varepsilon] \leq \frac{E[(X-\mu)^2]}{\varepsilon} .$$

Por tanto,

$$p[|X-\mu| \geq \sqrt{\varepsilon}] \leq \frac{\sigma^2}{\varepsilon} .$$

Tomando $\sqrt{\varepsilon} = K$, obtenemos:

$$p[|X-\mu| \geq K] \leq \frac{\sigma^2}{K^2} ,$$

como queríamos demostrar. ♣

Calculando la probabilidad del suceso complementario, podemos dar otra expresión de esta desigualdad:

$$p[|X-\mu| < K] \geq 1 - \frac{\sigma^2}{K^2} .$$

Mediante la desigualdad de CHEBYSHEV obtenemos una cota superior de la probabilidad de las *colas* de una distribución, es decir, una cota superior del área rayada en la figura 6.12.

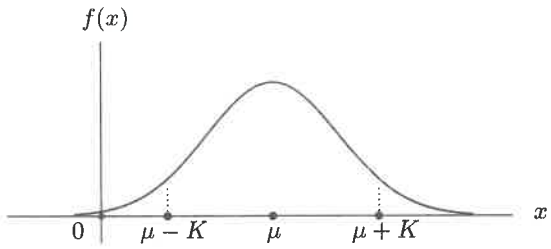


FIG. 6.12.

Ejemplo 6.17. Sea X el *número de clientes que acuden diariamente a un cierto establecimiento.* Sabiendo, únicamente, que el número medio de clientes es 200 y la

desviación 25, hallemos la probabilidad de que un día determinado acudan al menos 250 personas.

Al no conocer completamente la distribución de X, no podemos calcular dicha probabilidad con exactitud. Utilizando la desigualdad de CHEBYSHEV, y suponiendo que la distribución es más o menos simétrica, podremos obtener una cota superior de esta probabilidad. Como en este caso $p[X \geq 250] \approx p[X \leq 150]$, entonces,

$$\begin{aligned} p[X \geq 250] &\approx \frac{p[X \geq 250] + p[X \leq 150]}{2} = \\ &= \frac{1}{2} \cdot p[|X - 200| \geq 50] \leq \\ &\leq \frac{1}{2} \cdot \frac{\sigma^2}{50^2} = \frac{1}{8} . \end{aligned}$$

Por consiguiente, la probabilidad de que acudan al menos 250 clientes no será superior a $1/8$. ◁

Si en la desigualdad de CHEBYSHEV, damos a K los valores σ, 2σ, 3σ, ..., obtenemos:

$$\begin{aligned} &p[|X - \mu| \geq \sigma] \leq 1 , \\ &p[|X - \mu| \geq 2\sigma] \leq \frac{1}{4} , \\ &p[|X - \mu| \geq 3\sigma] \leq \frac{1}{9} , \\ &\ldots\ldots\ldots\ldots\ldots\ldots , \end{aligned}$$

es decir, el conjunto de valores de la variable que difieren de la esperanza, μ, en al menos 2, 3, 4, ... desviaciones típicas, tiene una probabilidad menor o igual que $\frac{1}{4}$, $\frac{1}{9}$, $\frac{1}{16}$, ..., respectivamente. En general, para cualquier valor δ mayor que cero,

$$p[|X - \mu| \geq \delta \cdot \sigma] \leq \frac{1}{\delta^2} .$$

Como hemos visto, la desigualdad de CHEBYSHEV permite calcular, por exceso, la probabilidad del suceso $[|X - \mu| \geq \varepsilon]$, o, por defecto, la de su complementario. La *holgura* del exceso o defecto correspondiente, dependerá tanto de la forma de la distribución, desconocida, como del valor de K.

Ejemplo 6.18. Sea X una variable aleatoria continua cuya función de densidad es:

$$f_X(x) = \begin{cases} 1 - |x| & \text{si } |x| < 1 \\ 0 & \text{en el resto} \end{cases}$$

Calculemos su esperanza y varianza:

$$E(X) = \int_{-1}^{0} x \cdot (1 + x)dx + \int_{0}^{1} x \cdot (1 - x)dx = 0$$

y, por tanto,

$$Var(X) = E(X^2) = \int_{-1}^{0} x^2 \cdot (1 + x)dx + \int_{0}^{1} x^2 \cdot (1 - x)dx = \frac{1}{6} .$$

Hallemos la probabilidad de que la variable difiera de la media al menos K, es decir, $p[|X-\mu| \geq K]$.

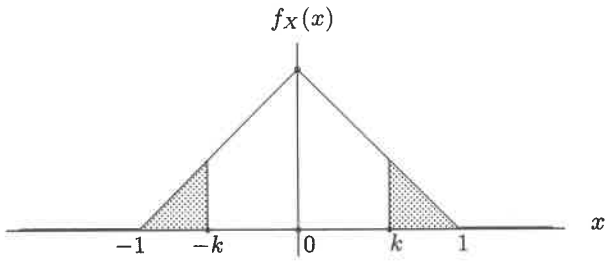


FIG. 6.13.

Como conocemos la distribución de X, podemos obtener esta probabilidad exactamente. Así, siendo $0 < K < 1$,

$$\begin{aligned} p[|X-0| \geq K] &= 1 - p[-K < X < K] = \\ &= 1 - \int_{-K}^{K} f_X(x)dx = (1-K)^2 . \end{aligned}$$

Puede el lector realizar el cálculo anterior valiéndose, únicamente, de procedimientos geométricos.

Supongamos que desconociésemos la distribución de X, sabiendo solamente que μ es igual a cero y σ^2 es $\frac{1}{6}$. Entonces, podríamos calcular de forma aproximada la probabilidad anterior. Así,

$$p[|X-0| \geq K] \leq \frac{\sigma^2}{K^2} = \frac{1}{6K^2} .$$

Para $K = 0.5$,

$$p[|X-0| \geq 0.5] = 0.25 ,$$

mientras que, por la desigualdad de CHEBYSHEV,

$$p[|X-0| \geq 0.5] \leq 0.66 ,$$

es decir, la aproximación es mala, con una holgura de 0.41. En cambio, para $K = 0.9$, la probabilidad exacta y la cota superior son, respectivamente, 0.01 y 0.21, con una holgura de 0.2. ◁

6.5 Ejercicios propuestos

6.1. Un individuo con un capital de medio millón de pesetas, desea invertirlo en un negocio cuya rentabilidad es del 50 %, pero con el posible riesgo de perder el total de la inversión. Su asesor financiero le ha informado que este negocio tiene una probabilidad de 0.8 de ser rentable. Si decide invertir, ¿cuál será su beneficio esperado?

6.2. Debido a los graves problemas económicos que atraviesa un cierto país, el gobierno ha decidido que los visitantes de uno de los parques naturales deberán pagar entrada a partir del próximo verano.
Se ha estimado que la variable X, número de personas por coche que entran en el parque, tiene la siguiente función de probabilidad:

X	1	2	3	4	5
$p[X=x]$	0.15	0.20	0.35	0.20	0.10

a) Hállese el número medio de visitantes por vehículo.
b) ¿Cuánto debe pagar cada persona para que la ganancia esperada por coche sea de 290 pesetas?
c) Si cada visitante paga P pesetas, ¿cuál es la ganancia esperada de un día en el que se registre una entrada de 1000 vehículos?

6.3. En Estadística se dice que un juego es justo cuando la ganancia esperada de cada jugador es cero.
Dos jugadores, A y B, tiran un dado por turnos, ganando aquel que obtenga primero un uno. Cada jugador apuesta una cantidad c_i $(i=1,2)$ cuyo total será finalmente para el ganador. Suponiendo que empieza el jugador A, ¿qué condición deben cumplir c_1 y c_2 para que el juego sea justo?

6.4. El rendimiento, en toneladas, de una hectárea de trigo es una variable aleatoria, X, que se modeliza según la siguiente función de densidad:

$$f(x)=\begin{cases} 3\frac{(x-1)(3-x)}{4} & 1\le x\le 3 \\ 0 & \text{en el resto} \end{cases}$$

Sabiendo que el agricultor percibe 20 pesetas por kilo de trigo y una cantidad fija de 5000 pesetas por hectárea cultivada, hállese:

a) La media y la varianza de la variable X.
b) El ingreso medio por hectárea.

6.5. El tiempo, en días, que un cliente tarda en abonar una factura en una tienda de electrodomésticos, es una variable aleatoria con función de distribución:

$$F(x)=\begin{cases} 1-\frac{2}{3}e^{-x/2} & \text{si } x\ge 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese el porcentaje de clientes que pagan al contado.
b) Calcúlese el tiempo medio de pago.

6.6. Una empresa constructora se compromete a realizar un edificio en 150 días, recibiendo por ello 100 millones de pesetas. Por cada semana de retraso en la entrega de la obra, la constructora deberá pagar una indemnización de 2 millones de pesetas. Se estima que el tiempo, en días, que se tarda en construir el edificio es una variable aleatoria cuya función de distribución es la siguiente:

$$F_X(x)=\begin{cases} 0 & \text{si } x<150 \\ 1-\frac{1}{3}e^{-(x-150)} & \text{si } x\ge 150 \end{cases}$$

a) ¿En cuántos días se espera concluir la obra?

b) ¿Qué cantidad espera recibir la constructora por la realización de dicha obra?

6.7. Sea X una variable aleatoria discreta que toma como valores los números naturales $1, 2, \ldots$

a) Demuéstrese que, si existe, $E(X) \geq 1$.

b) Compruébese que, si $E(X) = 1$, entonces, $p[X = 1] = 1$.

6.8. Sea X una variable aleatoria discreta que toma los valores $1, 2, \ldots$ Demuéstrese que, cuando exista, $E(X) = \sum_{k=1}^{\infty} p[X \geq k]$.

6.9. Sea X una variable aleatoria negativa. Pruébese que, si existe $E(X)$, ésta siempre es menor o igual que 0.

6.10. El precio por estacionamiento en un parking es de 75 pesetas para la primera hora o fracción, siendo de 60 pesetas a partir de la segunda hora o fracción. Se supone que el tiempo, en horas, que un vehículo permanece en el parking, se modeliza según la función de densidad

$$f(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

Hállese el ingreso medio por aparcamiento.

6.11. La compañía Petroleum Pluf, para satisfacer la demanda mensual de petróleo de la República de Ruritania, ofrece dos opciones:
Opción 1. El país deberá pagar 20 dólares por cada barril.
Opción 2. Si la demanda mensual no supera la cota de medio millón de barriles, el país deberá pagar una cantidad fija de 15 millones de dólares. Si la demanda mensual supera el medio millón de barriles, deberá pagar, además de los 15 millones fijos, 14 dólares por cada barril que supere dicha cota.
Se ha estimado, para Ruritania, que la demanda mensual de petróleo, en millones de barriles, tiene la siguiente función de densidad:

$$f_X(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 < x \leq 2 \\ 0 & \text{en el resto} \end{cases}$$

Sea Y_i $(i = 1, 2)$ la cantidad mensual, en millones de dólares, que deberá pagar el país, si se decide por la opción i-ésima.

a) Exprésense las variables Y_1 e Y_2 en función de X.

b) ¿Qué opción será más favorable al país?

c) El presidente de Ruritania, aconsejado por su primer ministro, se decide por la segunda opción. Desconfiando de él, le pone a prueba: durante el próximo año, si al menos en la mitad de los meses la cantidad a pagar por el país es superior a la que hubiera tenido que pagar de haber elegido la otra opción, el presidente mandará ajusticiar al primer ministro. Hállese la probabilidad de que esto ocurra.

6.12. Sea X una variable aleatoria discreta y sea $Y = g(X)$ una variable aleatoria función de X. Demuéstrese que, si existen,

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p[X = x_i]$$

y

$$E(Y) = \sum_{j=1}^{\infty} y_j p[Y = y_j]$$

coinciden.

6.13. Sea X una variable aleatoria continua y sea $Y = g(X)$ una función de X, donde g verifica las condiciones de la proposición 4.14. Demuéstrese que, si existen,

$$E[g(X)] = \int_{x=-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

y

$$E(Y) = \int_{y=-\infty}^{\infty} y \cdot f_Y(y)dy$$

coinciden.

6.14. Sea X una variable aleatoria cuya función de densidad es

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Obténgase la distribución de $Y = \ln(X)$ y la de $Z = \frac{1}{X}$.

b) Calcúlese la esperanza de la variable Y.

c) Compruébese que no existe la esperanza de la variable Z.

6.15. Pruébese que, si X es una variable aleatoria simétrica respecto de a, la variable $Y = X - a$ es simétrica respecto de cero.

6.16. Hállese la expresión del momento factorial de orden k como combinación lineal de los momentos no centrales de orden igual o inferior a k.

6.17. Sea X una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ a(1+x) & 0 < x \leq 1 \\ \frac{2}{3} & 1 < x \leq 2 \\ 0 & x > 2 \end{cases}$$

Calcúlese:

a) El valor de a para que $f_X(x)$ sea función de densidad.

b) La esperanza y la varianza de X.

6.18. El salario de los trabajadores de una empresa es una variable aleatoria con media 80 mil pesetas y desviación típica 20 mil. La patronal propone que la cantidad total correspondiente a la subida de los sueldos del próximo año se reparta entre los empleados, de modo que todos los salarios se incrementen en idéntico porcentaje. Frente a esta alternativa, y con el fin de disminuir las diferencias salariales, los sindicatos consideran oportuno que dicho total se distribuya de forma que cada uno de los empleados vea incrementado su salario en una misma cantidad. Justifíquese estadísticamente la propuesta sindical.

6.19. La altura de los niños que nacen al mes en el servicio de maternidad de un hospital comarcal es una variable aleatoria con esperanza 45 centímetros y desviación 3 centímetros, siendo estos valores 3.5 y 0.4 kilogramos, respectivamente, para la variable peso al nacer.
Hoy ha nacido una niña con un peso de 4 kilogramos y una altura de 60 centímetros. ¿Cuál de los dos valores es relativamente mayor con respecto al resto de los bebés que nacen en el hospital?

6.20. La puntuación que un candidato obtiene en el primer ejercicio de una oposición es una variable aleatoria con la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{x}{20} & 0 \leq x < 4 \\ \frac{(1-x)/10}{3} & 4 \leq x \leq 10 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese la media y la varianza del número de puntos obtenidos en el primer examen, así como una medida relativa del grado de dispersión de dicha variable.

Por lo que respecta al segundo ejercicio, la distribución de la puntuación responde al siguiente modelo:

$$f(y) = \begin{cases} \frac{x}{25} & 0 \leq x < 5 \\ \frac{10-x}{25} & 5 \leq x \leq 10 \\ 0 & \text{en el resto} \end{cases}$$

b) ¿En cuál de los dos exámenes se observa una mayor homogeneidad en las puntuaciones?

c) Un opositor obtiene unas calificaciones de 6 y 7 para el primer y segundo ejercicio, respectivamente. ¿En cuál de los dos ha conseguido un mejor resultado en relación con la actuación general de los opositores?

6.21. El salario mensual, en cientos de miles de pesetas, de los trabajadores de una pequeña empresa, es una variable aleatoria con función de densidad:

$$f(x) = \begin{cases} x-1 & 1 \leq x < 2 \\ 3-x & 2 \leq x \leq 3 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese, analítica y gráficamente, el salario mínimo del 50 % de los trabajadores que más cobran.

b) ¿Cuál es el salario máximo del 10 % de los trabajadores que menos cobran?

6.22. El retraso en minutos que un autobús tiene al llegar a una determinada parada, es una variable aleatoria X que se modeliza según la siguiente función de densidad:

$$f(x) = \begin{cases} \frac{1}{b} & 0 < x < b \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese la mediana y el cuantil de orden q en función de b.

b) Sabiendo que en la mitad de las ocasiones sufre un retraso inferior a 3 minutos, calcúlese la mediana y el recorrido intercuartílico.

6.23. Sea X una variable aleatoria cuya función de distribución es

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{2}{3}x & 0 < x < \frac{1}{2} \\ \frac{2}{3}x + \frac{1}{3} & \frac{1}{2} \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Calcúlese la esperanza, la mediana y la varianza de X.

6.24. Sea X una variable aleatoria continua y simétrica. Demuéstrese que la media y la mediana coinciden.

6.25. Estúdiese el comportamiento de la moda y de la mediana ante cambios de origen y de escala.

6.26. En una perfumería se venden diariamente X unidades de un cierto perfume. La función de probabilidad de esta variable es:

X	0	1	2	3
$p[X = x]$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{1}{7}$	$\frac{1}{7}$

a) Hállese la media, la mediana y la moda de esta distribución.

b) Obténgase la varianza, la desviación típica, el coeficiente de variación, la desviación absoluta media con respecto a la mediana y el recorrido intercuartílico.

c) Calcúlense los coeficientes de asimetría y de curtosis.

6.27. Sea X una variable aleatoria con función de densidad,

$$f(x) = \frac{1}{2} \cdot e^{-|x|} \quad -\infty < x < \infty \, .$$

a) Calcúlese la media, la mediana y la moda de esta distribución.

b) Obténgase la desviación absoluta media respecto a la mediana y el recorrido intercuartílico.

c) Hállense los coeficientes de asimetría y de curtosis.

6.28. Dada la función:

$$f(x) = \begin{cases} \frac{2}{x^2} & 1 \leq x \leq 2 \\ 0 & \text{en el resto} \end{cases}$$

a) Calcúlese la mediana, el primer cuartil y el segundo decil de la distribución.

b) Hállense los coeficientes de asimetría y de curtosis.

6.29. Demuéstrese que los coeficientes de asimetría y de curtosis son invariantes ante cambios de origen y de escala.

6.30. Sea X una variable aleatoria discreta que toma los valores $-k$, 0 y k con probabilidades $\frac{1}{8}$, $\frac{3}{4}$ y $\frac{1}{8}$, respectivamente.

a) Hállese $p[|X - \mu| \geq 2 \cdot \sigma]$.

b) Obténgase una cota superior de la probabilidad anterior, utilizando la desigualdad de CHEBYSHEV.

6.31. El dueño de una zapatería está cansado de oír a su mujer que en la zapatería de enfrente acuden más clientes. Convencido de que, aunque así sea algún día, él tiene una afluencia de personas mucho más regular, decide hacérselo ver a su esposa, por lo que contrata los servicios de un estadístico. Éste, tras analizar las distribuciones del número de clientes diarios que entran en cada uno de los establecimientos, informa al zapatero que las medias y varianzas son, respectivamente, 100 y 81 para el comercio de la competencia y 95 y 25 para el suyo.

a) ¿Confirman estas deducciones la intuición del vendedor?

b) Hállese una cota superior del porcentaje de días en que el número de clientes que acuden a la zapatería es mayor de 150.

6.32. El número de periódicos vendidos diariamente en un quiosco es una variable aleatoria de media 200 y desviación típica 10. ¿Cuántos ejemplares diarios debe encargar el propietario del quiosco para que tenga una seguridad de al menos un 99 % de no quedarse sin existencias?

6.33. El ayuntamiento de una ciudad ha decidido establecer una zona de estacionamiento limitado, en la que aparcan 500 vehículos por término medio. Sabiendo que al menos el 99 % de los días el número de coches que utilizan este servicio está entre 475 y 525, hállese una medida del grado de dispersión de la variable X, número de vehículos diarios que ocupan plazas de estacionamiento limitado.

6.34. El número de días al año que un trabajador de un pequeño comercio está de baja por enfermedad es una variable aleatoria de la que, únicamente, se conoce su media, 10, y su desviación típica, 2. Si cada uno de estos días supone a la empresa unas pérdidas de 10 mil pesetas, determínense los límites superior e inferior de las pérdidas anuales por trabajador con un grado de fiabilidad de al menos un 95 %.

Capítulo 7

CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA n-DIMENSIONAL

Comenzaremos este capítulo estudiando las características de las variables aleatorias n-dimensionales que, como indicamos en el caso unidimensional, son resúmenes de distintos aspectos de su distribución conjunta de probabilidades. Más adelante, utilizaremos estas características para evaluar el tipo y el grado de relación entre las variables.

El estudio de las características de distribuciones n-variantes nos conducirá a una de las materias más atractivas de la Estadística, por cuanto su campo de aplicación, tanto en las ciencias sociales como en las ciencias de la naturaleza, es muy amplio y sugerente. Se trata de encontrar relaciones funcionales entre variables aleatorias, que permitan explicar una de ellas en función de otras y, en último término, predecir su valor.

Así, podremos explicar (y predecir) el consumo de una familia en función de su renta y su tamaño, o bien las ventas de una empresa en función de su inversión publicitaria, su esfuerzo tecnológico y su capital fijo, o la altura de los hijos en función de la altura de sus padres. Este último problema está en el origen de las técnicas que vamos a estudiar, y que se conocen con el nombre de regresión.

En este capítulo haremos un desarrollo general de esta materia. Su importancia en sí misma y la relevancia de sus aplicaciones económicas, han provocado el nacimiento de una rama científica, desgajada hoy en día parcialmente de la Estadística, y que constituye una herramienta básica en la investigación económica: la Econometría.

7.1 Momentos de una variable aleatoria n-dimensional

Al igual que hicimos en el capítulo anterior con las variables aleatorias unidimensionales, vamos a definir los momentos de una variable bidimensional. El estudio se realiza únicamente para los casos en que dicha variable sea discreta o continua. Asimismo, generalizaremos las definiciones y resultados obtenidos a variables n-dimensionales.

Definición 7.1 *Sea (X,Y) una variable aleatoria bidimensional y g una aplicación de $\mathbb{R}^2$ en $\mathbb{R}$ tal que $g(X,Y)$ es una variable aleatoria unidimensional. Llamamos esperanza de $g(X,Y)$ al valor*

$$E[g(X,Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)\cdot f_{(X,Y)}(x_i,y_j) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)\cdot p[X=x_i, Y=y_j]\ ,$$

si (X,Y) es discreta, y

$$E[g(X,Y)] = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} g(x,y)\cdot f_{(X,Y)}(x,y)\,dxdy\ ,$$

si (X,Y) es continua, donde $f_{(X,Y)}$ es la función de probabilidad o de densidad, según el caso.

La esperanza de una función $g(X,Y)$ existe siempre que $E[|g(X,Y)|]$ sea finita.

Téngase en cuenta que la cantidad que hemos definido no es la esperanza de una variable bidimensional, sino la esperanza de una variable unidimensional función de aquélla. Por ello, su cálculo se podría realizar, también, hallando previamente la función de densidad de la variable $Z = g(X,Y)$, mediante un cambio de variable, y aplicando las definiciones 6.1 y 6.2 correspondientes al cálculo de la esperanza de una variable aleatoria unidimensional. En concreto, si Z es discreta con función de probabilidad f_Z, dicha esperanza sería

$$E[g(X,Y)] = E(Z) = \sum_{h=1}^{\infty} z_h\cdot f_Z(z_h) = \sum_{h=1}^{\infty} z_h\cdot p[Z=z_h]\ ,$$

y, si Z es continua con función de densidad f_Z,

$$E[g(X,Y)] = E(Z) = \int_{z=-\infty}^{\infty} z\cdot f_Z(z)dz\ .$$

No demostraremos que dichas técnicas de cálculo suministran el mismo valor numérico; dicha demostración no entraña dificultad, puesto que se trata, únicamente, de un cambio de variable; sin embargo, la omitiremos por lo engorroso de su desarrollo. El siguiente ejemplo ilustra este resultado.

Ejemplo 7.1. Se lanzan dos dados perfectos de forma independiente. Sean X e Y las variables aleatorias que recogen el resultado de cada uno de los dos lanzamientos. Vamos a hallar la suma esperada de las puntuaciones, es decir, $E(X+Y)$.

La distribución conjunta es sencilla de obtener, pues teniendo en cuenta que X e Y son independientes,

$$p[X=x, Y=y] = p[X=x]\cdot p[Y=y] = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}\ ,$$

para $x, y = 1,\ldots,6$.

La esperanza de la suma, $X+Y$, es

$$\begin{aligned} E(X+Y) &= \sum_{x=1}^{6}\sum_{y=1}^{6}(x+y)\cdot p[X=x, Y=y] = \\ &= (1+1)\cdot\frac{1}{36}+\cdots+(6+6)\cdot\frac{1}{36} = 7 \text{ puntos.} \end{aligned}$$

También podríamos haber obtenido, previamente, la distribución de la suma, $Z = X+Y$. La siguiente tabla describe su función de probabilidad:

Z	2	3	4	5	6	7	8	9	10	11	12
$p[Z=z]$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Entonces,

$$E(X+Y) = E(Z) = 2\cdot\frac{1}{36}+\cdots+12\cdot\frac{1}{36} = 7 \text{ puntos.}$$

Como se ve, las dos técnicas conducen al mismo resultado. ◁

La anterior definición 7.1 se puede generalizar a una variable aleatoria n-dimensional $(X_1,\ldots,X_n)$ discreta o continua.

Definición 7.2 *Sea $(X_1,\ldots,X_n)$ una variable aleatoria n-dimensional y g una aplicación de $\mathbb{R}^n$ en $\mathbb{R}$ tal que $g(X_1,\ldots,X_n)$ es una variable aleatoria unidimensional. Llamamos esperanza de $g(X_1,\ldots,X_n)$ al valor*

$$\begin{aligned} E[g(X_1,\ldots,X_n)] &= \sum_{x_1}\cdots\sum_{x_n} g(x_1,\ldots,x_n)\cdot f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \\ &= \sum_{x_1}\cdots\sum_{x_n} g(x_1,\ldots,x_n)\cdot p[X_1=x_1,\ldots,X_n=x_n]\,, \end{aligned}$$

si la variable n-dimensional es discreta, y

$$E[g(X_1,\ldots,X_n)] = \int_{x_1=-\infty}^{\infty}\cdots\int_{x_n=-\infty}^{\infty} g(x_1,\ldots,x_n)\cdot f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n)dx_1\ldots dx_n\,,$$

si $(X_1,\ldots,X_n)$ es continua, donde $f_{(X_1,\ldots,X_n)}$ es la función de probabilidad o de densidad, según el caso.

Adviértase que en la definición de $E[g(X_1,\ldots,X_n)]$, los sumatorios que aparecen para el caso discreto, se extienden a todo el campo de variación de las variables.

La esperanza de $g(X_1,\ldots,X_n)$ existe si $E[|g(X_1,\ldots,X_n)|]$ es finita.

Veamos, a continuación, algunas propiedades, que enunciaremos para variables aleatorias n-dimensionales, pero que, por comodidad, demostraremos para bidimensionales, suponiendo en todos los casos que no hay problemas de existencia en las esperanzas.

P1. La esperanza de una combinación lineal de variables aleatorias es la combinación lineal de sus esperanzas; es decir, para $X_1, \ldots, X_n$, n variables aleatorias, y $a_1, \ldots, a_n$, n números reales cualesquiera,

$$E(a_1 \cdot X_1 + \cdots + a_n \cdot X_n) = a_1 \cdot E(X_1) + \cdots + a_n \cdot E(X_n) .$$

Demostración. Sea (X, Y) una variable aleatoria bidimensional continua con función de densidad $f_{(X,Y)}$. Entonces,

$$\begin{aligned} E(a \cdot X + b \cdot Y) &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} (a \cdot x + b \cdot y) \cdot f_{(X,Y)}(x,y) dx dy = \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} a \cdot x \cdot f_{(X,Y)}(x,y) dx dy + \\ &\quad + \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} b \cdot y \cdot f_{(X,Y)}(x,y) dx dy = \\ &= \int_{x=-\infty}^{\infty} a \cdot x \cdot \left[\int_{y=-\infty}^{\infty} f_{(X,Y)}(x,y) dy \right] dx + \\ &\quad + \int_{y=-\infty}^{\infty} b \cdot y \cdot \left[\int_{x=-\infty}^{\infty} f_{(X,Y)}(x,y) dx \right] dy = \\ &= a \cdot \int_{x=-\infty}^{\infty} x \cdot f_X(x) dx + b \cdot \int_{y=-\infty}^{\infty} y \cdot f_Y(y) dy = \\ &= a \cdot E(X) + b \cdot E(Y) . \end{aligned}$$

La demostración para variables discretas es análoga, sin más que sustituir las integrales por sumas. ♣

De la misma forma se prueba que, si $g_1, \ldots, g_k$ son k funciones de $\mathbb{R}^n$ en $\mathbb{R}$,

$$\begin{aligned} &E[a_1 \cdot g_1(X_1, \ldots, X_n) + \cdots + a_k \cdot g_k(X_1, \ldots, X_n)] = \\ &= a_1 \cdot E[g_1(X_1, \ldots, X_n)] + \cdots + a_k \cdot E[g_k(X_1, \ldots, X_n)] . \end{aligned}$$

Como consecuencias de la propiedad anterior, se deducen dos importantes corolarios.

a) La esperanza de una suma de variables aleatorias es la suma de las esperanzas de cada una de ellas. Es decir,

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) .$$

Demostración. Este resultado se obtiene del anterior, tomando $a_1 = \cdots = a_n = 1$. ♣

El siguiente corolario se refiere a una variable aleatoria de especial interés: la media muestral. *Dadas n variables aleatorias $X_1, \ldots, X_n$, su media muestral se define como*

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n} .$$

b) Sean $X_1, \ldots, X_n$ variables aleatorias con la misma esperanza μ. Entonces, la esperanza de la media muestral es la esperanza de las variables, es decir,

$$E(\overline{X}) = \mu \, .$$

Demostración. Dado que $\overline{X}$ es una combinación lineal de $X_1, \ldots, X_n$ cuyos coeficientes valen $a_1 = \cdots = a_n = \frac{1}{n}$,

$$E(\overline{X}) = E\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{E(X_1) + \cdots + E(X_n)}{n} = \frac{n \cdot \mu}{n} = \mu \, . \clubsuit$$

P2. La esperanza del producto de variables aleatorias independientes es el producto de sus esperanzas, esto es,

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n) \, .$$

Demostración. Sea (X, Y) una variable bidimensional continua. Si las variables son independientes, la función de densidad conjunta coincide con el producto de las marginales, es decir

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y) \, .$$

Por tanto,

$$\begin{aligned} E(X \cdot Y) &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} x \cdot y \cdot f_{(X,Y)}(x, y) dx dy = \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} x \cdot y \cdot f_X(x) \cdot f_Y(y) dx dy = \\ &= \int_{x=-\infty}^{\infty} x \cdot f_X(x) dx \cdot \int_{y=-\infty}^{\infty} y \cdot f_Y(y) dy = E(X) \cdot E(Y) \, . \end{aligned}$$

La demostración para variables discretas y su generalización para el caso n-dimensional no presentan ninguna dificultad. $\clubsuit$

La propiedad recíproca no es cierta, como pone de manifiesto el siguiente ejemplo.

Ejemplo 7.2. La siguiente tabla de doble entrada recoge la distribución conjunta de una variable aleatoria bidimensional discreta y de sus marginales.

$X \backslash Y$	0	1	
-1	0.1	0.2	0.3
0	0.3	0.1	0.4
1	0.1	0.2	0.3
	0.5	0.5	

Fácilmente se comprueba que X e Y son dependientes; por ejemplo,

$$p[X = -1, Y = 0] \neq p[X = -1] \cdot p[Y = 0] \, .$$

Calculemos las esperanzas de X e Y y la de su producto:

$$\begin{aligned} E(X) &= -1 \cdot 0.3 + 0 \cdot 0.4 + 1 \cdot 0.3 = 0 \\ E(Y) &= 0 \cdot 0.5 + 1 \cdot 0.5 = 0.5 \, . \end{aligned}$$

La distribución de la variable $X \cdot Y$ es:

$$X \cdot Y : \left| \begin{array}{ll} -1 & p[X \cdot Y = -1] = 0.2 \\ 0 & p[X \cdot Y = 0] = 0.6 \\ 1 & p[X \cdot Y = 1] = 0.2 \end{array} \right.$$

y, por tanto, $E(X \cdot Y) = 0$, con lo que, evidentemente,

$$E(X \cdot Y) = E(X) \cdot E(Y) \, .$$

Es decir, la esperanza del producto es el producto de las esperanzas, siendo X e Y dependientes. Consecuentemente, el recíproco de la propiedad **2** no es cierto. ◁

Para las variables aleatorias unidimensionales hemos definido una serie de momentos que resumen ciertos aspectos de su distribución. Para las variables aleatorias bidimensionales y, en general, para las n-dimensionales, se pueden definir unos momentos de forma similar. Desarrollaremos solamente los momentos centrados respecto del origen y respecto de la media para variables bidimensionales.

Definición 7.3 *Sea (X, Y) una variable aleatoria bidimensional. El momento centrado respecto del origen de orden r en X y de orden s en Y o momento no central, es el valor, si existe,*

$$\alpha_{r,s} = E(X^r \cdot Y^s) \qquad r, s = 0, 1, 2, \ldots$$

Haciendo, o bien $s = 0$, o bien $r = 0$, obtendríamos los correspondientes momentos centrados respecto del origen de X y de Y. Entre ellos, los que se utilizan más frecuentemente son

$$\alpha_{1,0} = E(X) \, , \quad \alpha_{0,1} = E(Y) \, , \quad \alpha_{2,0} = E(X^2) \quad \text{y} \quad \alpha_{0,2} = E(Y^2) \, .$$

Además, en el estudio de la relación entre X e Y, tendremos que calcular el momento

$$\alpha_{1,1} = E(X \cdot Y) \, .$$

Definición 7.4 *Sea (X, Y) una variable aleatoria bidimensional cuyas esperanzas designaremos por μ_X y μ_Y, respectivamente. El momento centrado respecto de las medias de orden r en X y de orden s en Y o momento central, es el valor, si existe,*

$$\mu_{r,s} = E[(X - \mu_X)^r \cdot (Y - \mu_Y)^s] \qquad r, s = 0, 1, 2, \ldots$$

Anulando s o r, obtenemos los correspondientes momentos centrales de X y de Y, respectivamente. Los más utilizados son los de orden dos,

$$\mu_{2,0} = E[(X - \mu_X)^2] = Var(X) \qquad \text{y} \qquad \mu_{0,2} = E[(Y - \mu_Y)^2] = Var(Y) \, .$$

Además, es de especial interés el momento

$$\mu_{1,1} = E[(X - \mu_X) \cdot (Y - \mu_Y)] ,$$

que denominaremos **covarianza** entre X e Y, y que designaremos, también, como $Cov(X,Y)$, μ_{XY} o σ_{XY}.

De la definición de covarianza se deduce que si (X,Y) es una variable bidimensional discreta con función de probabilidad $f_{(X,Y)}$,

$$\begin{aligned} Cov(X,Y) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - \mu_X) \cdot (y_j - \mu_Y) \cdot f_{(X,Y)}(x_i, y_j) = \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - \mu_X) \cdot (y_j - \mu_Y) \cdot p[X = x_i, Y = y_j] , \end{aligned}$$

y si (X,Y) es continua con función de densidad $f_{(X,Y)}$, entonces,

$$Cov(X,Y) = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty}(x - \mu_X) \cdot (y - \mu_Y) \cdot f_{(X,Y)}(x,y)\, dxdy .$$

Como veremos más adelante, la covarianza va a permitir dar una medida para evaluar la relación lineal existente entre las variables X e Y.

Asimismo, de la definición de covarianza se deduce de inmediato que

$$Cov(X,Y) = Cov(Y,X)$$

y que

$$Cov(X,X) = Var(X) .$$

Ejemplo 7.3. Sea (X,Y) una variable aleatoria bidimensional continua cuya función de densidad conjunta es

$$f(x,y) = \begin{cases} 8xy & 0 \le y \le x \le 1 \\ 0 & \text{en el resto} \end{cases}$$

Vamos a hallar la covarianza entre X e Y, es decir,

$$Cov(X,Y) = E[(X - \mu_X) \cdot (Y - \mu_Y)] .$$

Calculemos, previamente, las medias μ_X y μ_Y:

$$\mu_X = \int_{x=0}^{1}\int_{y=0}^{x} x \cdot 8xy\, dydx = \frac{4}{5}$$

y

$$\mu_Y = \int_{x=0}^{1}\int_{y=0}^{x} y \cdot 8xy\, dydx = \frac{8}{15} .$$

Entonces,

$$Cov(X,Y) = \int_{x=0}^{1}\int_{y=0}^{x}\left(x - \frac{4}{5}\right) \cdot \left(y - \frac{8}{15}\right) \cdot 8xy\, dydx = \frac{4}{225} . \triangleleft$$

Estudiemos, ahora, algunas de las propiedades de la covarianza.

P1. La covarianza entre dos variables se puede escribir en función de los momentos no centrales. En concreto,

$$Cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) .$$

Demostración. Sin más que desarrollar la definición de covarianza,

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)\cdot(Y-\mu_Y)] = \\ &= E(X\cdot Y - X\cdot\mu_Y - \mu_X\cdot Y + \mu_X\cdot\mu_Y) = \\ &= E(X\cdot Y) - E(X)\cdot E(Y) . \clubsuit \end{aligned}$$

Esta propiedad, que podemos escribir como $\mu_{1,1} = \alpha_{1,1} - \alpha_{1,0}\cdot\alpha_{0,1}$, permitirá calcular covarianzas más fácilmente que utilizando la expresión que la define.

P2. Sean X e Y dos variables aleatorias, y a, b, c y d cuatro números reales cualesquiera. Sean las variables $U = a\cdot X + b$ y $V = c\cdot Y + d$. Entonces,

$$Cov(U,V) = a\cdot c\cdot Cov(X,Y) .$$

Demostración.

$$\begin{aligned} Cov(U,V) &= E[(U-E(U))(V-E(V))] = \\ &= E[(a\cdot X + b - a\cdot E(X) - b)(c\cdot Y + d - c\cdot E(Y) - d)] = \\ &= E[a(X-E(X))\cdot c(Y-E(Y))] = a\cdot c\cdot Cov(X,Y) . \clubsuit \end{aligned}$$

Un caso particular del doble cambio de variable anterior es un cambio de origen y de escala en cada una de las variables. Así, sean U y V las transformadas de X y de Y por un cambio de origen y de escala, es decir,

$$U = \frac{X - o_X}{e_X} \qquad \text{y} \qquad V = \frac{Y - o_Y}{e_Y} .$$

Entonces, la covarianza de las nuevas variables será:

$$\begin{aligned} Cov(U,V) &= Cov\left(\frac{X-o_X}{e_X}, \frac{Y-o_Y}{e_Y}\right) = \\ &= \frac{1}{e_X\cdot e_Y}\cdot Cov(X,Y) . \end{aligned}$$

Por tanto, los cambios de origen en las variables no afectan al valor de la covarianza, pero sí los cambios de escala.

P3. Sean $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_m$ dos colecciones disjuntas de variables aleatorias, y sean $a_1, \ldots, a_n$ y $b_1, \ldots, b_m$ dos colecciones de números reales cualesquiera. Entonces,

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j) .$$

Demostración. De acuerdo con la primera propiedad demostrada,

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = E\left(\sum_{i=1}^{n} a_i X_i \cdot \sum_{j=1}^{m} b_j Y_j\right) - E\left(\sum_{i=1}^{n} a_i X_i\right) \cdot E\left(\sum_{j=1}^{m} b_j Y_j\right) .$$

Desarrollando el producto de las combinaciones lineales de las variables, la expresión anterior resulta:

$$E\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j X_i Y_j\right) - \left[\sum_{i=1}^{n} a_i E(X_i)\right] \cdot \left[\sum_{j=1}^{m} b_j E(Y_j)\right] .$$

Haciendo lo mismo con las combinaciones lineales de sus esperanzas,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E(X_i \cdot Y_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E(X_i) \cdot E(Y_j) ,$$

obtenemos el resultado deseado:

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j) . \clubsuit$$

Obviamente, si las dos colecciones tienen variables comunes, los sumandos correspondientes incluirán la varianza en lugar de la covarianza. La siguiente propiedad recoge el caso extremo, cuando las dos colecciones de variables coinciden.

P4. Sean $X_1, \ldots, X_n$ una colección de variables aleatorias, y sean $a_1, \ldots, a_n$ y $b_1, \ldots, b_n$ dos colecciones de números reales cualesquiera. Entonces,

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} a_i b_i Var(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} a_i b_i Cov(X_i, X_j) .$$

Demostración.

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} b_i X_i\right) = E\left(\sum_{i=1}^{n} a_i X_i \cdot \sum_{i=1}^{n} b_i X_i\right) - E\left(\sum_{i=1}^{n} a_i X_i\right) \cdot E\left(\sum_{i=1}^{n} b_i X_i\right) ,$$

expresión que, multiplicando las combinaciones lineales de las variables, resulta:

$$E\left(\sum_{i=1}^{n} a_i b_i X_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} a_i b_j X_i X_j\right) - \left[\sum_{i=1}^{n} a_i E(X_i)\right] \cdot \left[\sum_{i=1}^{n} b_i E(X_i)\right] .$$

Repitiendo el procedimiento con las combinaciones lineales de las esperanzas,

$$\sum_{i=1}^{n} a_i b_i E(X_i^2) + \sum_{i=1}^{n} \sum_{\substack{i=j \\ i \neq j}}^{n} a_i b_j E(X_i \cdot X_j) - \sum_{i=1}^{n} a_i b_j [E(X_i)]^2 - \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i b_j E(X_i) E(X_j) ,$$

de donde se deduce que

$$Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} a_i b_i Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_i b_j Cov(X_i, X_j) . \clubsuit$$

P5. Si X e Y son variables aleatorias independientes, su covarianza es nula.

Demostración. Es inmediata, teniendo en cuenta la propiedad probada anteriormente que afirma que si X e Y son independientes, la esperanza de su producto coincide con el producto de sus esperanzas. Al igual que entonces, la propiedad recíproca no es cierta (el lector puede comprobarlo utilizando el mismo ejemplo 7.2). $\clubsuit$

P6. La varianza de la suma de dos variables es igual a la suma de sus varianzas más el doble de su covarianza; es decir,

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) .$$

Demostración.

$$\begin{aligned} Var(X+Y) &= E[(X+Y-E(X+Y))^2] = \\ &= E[((X-E(X)) + (Y-E(Y)))^2] , \end{aligned}$$

que, desarrollando, resulta

$$E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y))] ,$$

es decir,

$$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y) . \clubsuit$$

El resultado es fácilmente generalizable a la suma de n variables aleatorias:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} Cov(X_i, X_j) .$$

P7. La varianza de una suma de variables aleatorias independientes es igual a la suma de las varianzas. Es decir, si $X_1, \ldots, X_n$ son variables independientes,

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) .$$

La demostración es inmediata, aplicando las dos propiedades anteriores, ya que si las variables son independientes, son independientes dos a dos y, por tanto, la covarianza entre dos cualesquiera de ellas es igual a cero.

P8. Sean $X_1, \ldots, X_n$ variables aleatorias, y sean $a_1, \ldots, a_n$, n números reales cualesquiera. Entonces, la varianza de la combinación lineal, $\sum_{i=1}^{n} a_i X_i$, será

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j>i}}^{n} a_i a_j Cov(X_i, X_j) .$$

Demostración. Utilizando la propiedad anteriormente demostrada sobre el valor de la varianza de una suma de variables, obtenemos

$$\begin{aligned} Var\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} Var(a_i X_i) + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j>i}}^{n} Cov(a_i X_i, a_j X_j) = \\ &= \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j>i}}^{n} a_i a_j Cov(X_i, X_j) \ . \clubsuit \end{aligned}$$

Si la combinación lineal tiene término independiente, es decir, es de la forma $a_0 + a_1 X_1 + \cdots + a_n X_n$, el resultado es idéntico, pues el término constante no afecta a la varianza.

P9. Si $X_1, \ldots, X_n$ son variables aleatorias independientes, y $a_1, \ldots, a_n$, n números reales cualesquiera,

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) .$$

Demostración. Se deduce de la propiedad anterior, considerando que, si las variables son independientes, lo son dos a dos, y, por tanto, su covarianza es nula. ♣

De esta propiedad se deduce un corolario de gran interés en los desarrollos de la Inferencia estadística.

a) Si $X_1, \ldots, X_n$ son variables aleatorias independientes de media μ y varianza σ^2, la varianza de su media muestral, $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$, vale

$$Var(\overline{X}) = \frac{\sigma^2}{n} .$$

Demostración. Este resultado es un caso particular de la propiedad anterior cuando $a_1 = \cdots = a_n = \frac{1}{n}$, o lo que es lo mismo,

$$Var(\overline{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n} \ . \clubsuit$$

Téngase en cuenta que, aunque las propiedades séptima y novena, y el corolario de ésta, se enuncian para variables independientes, únicamente se exige que las covarianzas entre dichas variables se anulen.

La covarianza nos va a permitir definir una nueva característica, denominada coeficiente de correlación lineal. Como veremos más adelante, con este coeficiente podremos medir, de forma inmediata, el grado de relación lineal que existe entre dos variables.

Definición 7.5 *Sea (X,Y) una variable aleatoria bidimensional. Definimos el coeficiente de correlación lineal entre las variables X e Y, que designaremos por ρ_{XY}, como*

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\cdot Var(Y)}} .$$

Cuando no exista confusión lo denotaremos, simplemente, por ρ.

El coeficiente de correlación es, consecuentemente, el cociente entre la covarianza de X e Y y el producto de sus desviaciones típicas,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} .$$

Según veremos en la siguiente propiedad, el campo de variación es el intervalo $[-1,1]$; este hecho constituye una ventaja incuestionable frente a la covarianza, en el estudio del grado de relación lineal existente entre dos variables.

P1. El coeficiente de correlación lineal, ρ, está comprendido entre -1 y 1, es decir,

$$-1 \leq \rho \leq 1 .$$

Demostración. Desarrollemos la siguiente esperanza

$$\begin{aligned} E[(a(X-\mu_X)+b(Y-\mu_Y))^2] &= a^2Var(X)+b^2Var(Y)+2abCov(X,Y) = \\ &= (a,\ b)\begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} \geq 0 . \end{aligned}$$

Ahora bien, esta forma cuadrática es semidefinida positiva, al ser la esperanza de una variable al cuadrado; el determinante de la matriz tiene que ser, en consecuencia, positivo o nulo, esto es,

$$Var(X)\cdot Var(Y) - [Cov(X,Y)]^2 \geq 0 .$$

En definitiva,

$$Var(X)\cdot Var(Y) \geq [Cov(X,Y)]^2 ,$$

y, por tanto,

$$\rho^2 \leq 1 ,$$

o, lo que es lo mismo, ρ estará comprendido entre -1 y 1. ♣

Cuando el coeficiente de correlación entre dos variables X e Y es cero, se dice que estas variables están incorrelacionadas o que existe **incorrelación** *entre ellas.* Esto, obviamente, es equivalente al hecho de que la covarianza de X e Y sea cero.

Si el coeficiente de correlación es no nulo, diremos que las variables están correlacionadas. En este caso, si ρ es positivo, hablaremos de correlación positiva, y de correlación negativa si ρ es negativo. Posteriormente justificaremos estas denominaciones.

P2. Si $X_1, \ldots, X_n$ son variables aleatorias independientes, están incorrelacionadas.

Demostración. Si $X_1, \ldots, X_n$ son independientes, lo son dos a dos y, por consiguiente, la covarianza de dos cualesquiera de dichas variables se anula. En consecuencia, su coeficiente de correlación vale cero, es decir, están incorrelacionadas. ♣

El recíproco no es cierto: dos o más variables incorrelacionadas no tienen por qué ser independientes, como se confirma con el ejemplo 7.2. Por tanto, independencia implica incorrelación pero, en cambio, incorrelación no implica independencia.

Adviértase que ciertas propiedades enunciadas, y demostradas, en este mismo apartado para variables independientes, son igualmente válidas si únicamente exigimos incorrelación.

P3. Sean X e Y dos variables aleatorias, y a, b, c y d cuatro números reales cualesquiera. Sean las variables $U = a \cdot X + b$ y $V = c \cdot Y + d$. Entonces,

$$\rho_{UV} = \frac{a \cdot c}{|a| \cdot |c|} \cdot \rho_{XY} \ .$$

Demostración. Teniendo en cuenta que, por un lado,

$$Var(U) = Var(a \cdot X + b) = a^2 \cdot Var(X)$$

y

$$Var(V) = Var(c \cdot Y + d) = c^2 \cdot Var(Y) \ ,$$

y, por otro lado,

$$Cov(U, V) = Cov(a \cdot X + b, c \cdot Y + d) = a \cdot c \cdot Cov(X, Y) \ ,$$

obtenemos:

$$\rho_{UV} = \frac{\sigma_{UV}}{\sigma_U \cdot \sigma_V} = \frac{a \cdot c \cdot \sigma_{XY}}{|a| \cdot \sigma_X \cdot |c| \cdot \sigma_Y} = \frac{a \cdot c}{|a| \cdot |c|} \cdot \rho_{XY} \ . \ \clubsuit$$

Entonces, si a y c tienen el mismo signo,

$$\rho_{UV} = \rho_{XY} \ ,$$

y si a y c tienen signos distintos,

$$\rho_{UV} = -\rho_{XY} \ .$$

Puesto que los cambios de origen y de escala en X y en Y son un caso particular del cambio de variable considerado en la propiedad anterior, podemos afirmar que el coeficiente de correlación es invariante ante cualquier cambio de origen y de escala. Es decir, si

$$U = \frac{X - o_X}{e_X} \qquad \text{y} \qquad V = \frac{Y - o_Y}{e_Y} ,$$

el coeficiente de correlación de U y V coincide con el de X e Y:

$$\rho_{UV} = \rho_{XY} ,$$

ya que, según comentamos al introducir este tipo de cambios, e_X y e_Y son siempre números reales positivos.

P4. El coeficiente de correlación lineal entre las variables X e Y es igual a la covarianza de las variables X e Y tipificadas. Es decir, si definimos U y V como

$$U = \frac{X - \mu_X}{\sigma_X} \qquad \text{y} \qquad V = \frac{Y - \mu_Y}{\sigma_Y} ,$$

se cumple que

$$Cov(U, V) = \rho_{XY} .$$

Demostración. Debido al efecto que los cambios de origen y de escala producen en la covarianza,

$$Cov(U, V) = Cov\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y} = \rho_{XY} . \clubsuit$$

Con la ayuda de esta propiedad podremos interpretar el sentido de la covarianza y del coeficiente de correlación lineal. Cuando tipificamos un par de variables X e Y, la **nube de puntos**[1] se traslada de lugar en el plano, cambiando su centro, denominado centro de gravedad de la nube, al origen de coordenadas; además, la nube se concentra o expande (dependiendo de los valores de las varianzas de X e Y) respecto de cada uno de los ejes.

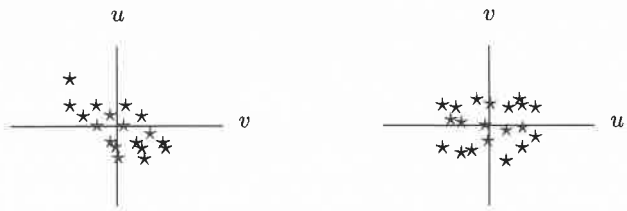
1. La idea de imaginar una distribución de probabilidades de una variable aleatoria bidimensional como una nube de puntos está tomada de la Estadística descriptiva. Cuando se estudia una variable estadística bidimensional, uno puede imaginar que sus valores, que son puntos del plano, llevan asociados pesos distintos dependiendo de su frecuencia (algo así como si unos puntos fuesen más *gruesos* que otros). Esa representación de una variable estadística se denomina **nube de puntos**. Cuando en un experimento aleatorio se dispone de una variable bidimensional discreta, uno puede considerar una nube de puntos de la misma forma, sin más que asociar como pesos de los puntos su probabilidad, en lugar de su frecuencia. Pero si como consecuencia de un experimento aleatorio, se generara una variable bidimensional continua, haría falta un notable esfuerzo de imaginación para considerar que los valores de la variable (que ahora podrían ser todos los del plano) forman una nube de puntos. Con todo, así se hace, y se acepta que cada punto lleva asociado un peso, que ahora se toma igual al valor de la función de densidad en el mismo.

FIG. 7.1.

Teniendo en cuenta las propiedades demostradas, y el hecho de que las variables tipificadas, U y V, son variables de media cero, obtenemos:

$$E(U \cdot V) = Cov(U, V) = \rho_{XY} \ .$$

Si los puntos de la nube tipificada se encuentran, principalmente, en el primer y tercer cuadrante del sistema de ejes cartesianos (como en la figura 7.1), $U \cdot V$ es una variable cuyos valores son mayoritariamente positivos (producto, bien de dos números positivos, bien de dos números negativos). Por tanto, $E(U \cdot V)$ será positiva, siéndolo, también, el coeficiente de correlación entre X e Y, es decir, las variables tendrán correlación positiva. Un valor positivo de ρ_{XY} pone en evidencia, por tanto, relaciones más o menos lineales, pero siempre crecientes, entre las variables.



a) b)

FIG. 7.2.

Si los puntos de la nube tipificada se encuentran principalmente en el segundo y cuarto cuadrante (figura 7.2.a), el razonamiento del párrafo anterior nos llevaría a la conclusión de que el coeficiente de correlación lineal entre X e Y sería negativo, esto es, existiría correlación negativa entre las variables. Este valor negativo de ρ_{XY} evidenciaría, así, relaciones más o menos lineales entre X e Y pero, en este caso, en sentido decreciente.

Por último, si los puntos de la nube se distribuyen de forma más o menos homogénea entre los cuatro cuadrantes (figura 7.2.b), los valores de $U \cdot V$ serán positivos y negativos, en proporciones más o menos iguales. De aquí se deduce

que $E(U \cdot V)$ tomará valores, al igual que ρ_{XY}, cercanos a cero. En consecuencia, relaciones no lineales entre las variables, dan lugar a coeficientes de correlación, positivos o negativos, cercanos a cero; es decir, podemos hablar, prácticamente, de ausencia de correlación, esto es, de ausencia de relación lineal entre las variables.

7.2 Esperanza condicionada

A veces, cuando un experimento aleatorio da lugar a la obtención de una variable aleatoria bidimensional (X, Y), se desea estudiar la distribución de una de ellas, X, condicionada por el conocimiento de que la otra, Y, ha tomado un valor concreto, o que su valor pertenece a un cierto conjunto. Para cada una de estas distribuciones condicionadas, nos podemos plantear el mismo tipo de resúmenes que viéramos en el capítulo anterior: medias, varianzas, medianas, etc. Esencialmente, las definiciones y propiedades que damos, en este apartado y en el siguiente, no son nuevas, sino que reproducen el cálculo de momentos estudiado en el capítulo anterior, al caso en que la distribución de probabilidades sea una distribución condicionada.

Definición 7.6 *Sea (X, Y) una variable aleatoria bidimensional y sea y un valor posible de Y (esto es, si la variable es discreta, $p[Y = y] \neq 0$, y si es continua, $f_Y(y) \neq 0$). Llamamos esperanza de X condicionada por $Y = y$ al valor, si existe,*

$$E(X \mid Y = y) = \sum_{i=1}^{\infty} x_i \cdot f_{(X \mid Y=y)}(x_i) = \sum_{i=1}^{\infty} x_i \cdot p[X = x_i \mid Y = y] \,,$$

si (X, Y) es discreta, y

$$E(X \mid Y = y) = \int_{x=-\infty}^{\infty} x \cdot f_{(X \mid Y=y)}(x)\, dx \,,$$

si (X, Y) es continua.

Análogamente, podríamos también considerar la esperanza de Y condicionada por $X = x$. Todos los desarrollos de esta sección y de la siguiente se realizarán a partir de distribuciones condicionadas por valores de Y, siendo igualmente válidos si condicionamos por valores de X.

OBSERVACIONES

O1. La esperanza de X condicionada por $Y = y$ existe si $E(|X| \mid Y = y)$ es finita. Los problemas de existencia de $E(X \mid Y = y)$ son consecuencia de los que se presenten al realizar las sumas o integrales que lo definen. Al igual que en el capítulo anterior, podemos afirmar que dichos problemas de existencia no se dan en las variables aleatorias de uso más frecuente.

O2. La idea intuitiva acerca del significado de la esperanza condicionada es evidente: es el valor esperado o medio de una de las variables cuando se conoce el valor de la otra.

O3. Para cada valor posible, y, de la variable aleatoria Y, existe una distribución condicionada de X y, por tanto, un valor $E(X \mid Y = y)$. La esperanza condicionada será, consecuentemente, una función en y,

$$E(X \mid Y = y) = g(y) .$$

O4. Para cualquier función h de X, $h(X)$, que sea variable aleatoria, resulta inmediata la generalización de la definición de esperanza condicionada, que, si existe, es

$$E[h(X) \mid Y = y] = \sum_{i=1}^{\infty} h(x_i) \cdot p[X = x_i \mid Y = y] ,$$

si (X, Y) *es discreta, o bien*

$$E[h(X) \mid Y = y] = \int_{x=-\infty}^{\infty} h(x) \cdot f_{(X \mid Y=y)}(x)\, dx ,$$

si (X, Y) *es continua.*

Más aún se utiliza, a veces, la esperanza de una función de X e Y condicionada por un valor de Y, y, esto es,

$$E[h(X, Y) \mid Y = y] .$$

En este caso,

$$E[h(X, Y) \mid Y = y] = E[h(X, y) \mid Y = y] ,$$

porque, conocido el valor, y, de Y, la función $h(X, Y)$ depende sólo de la aleatoriedad de la variable X, ya que la variable Y pierde su carácter aleatorio en el momento en que se conoce su valor.

O5. En ocasiones, se ignora el valor concreto que ha tomado la variable aleatoria Y tras realizar el experimento, y sólo se dispone de información parcial sobre el valor de Y obtenido. Podemos condicionar, entonces, la variable X por el conjunto de valores de Y que contengan dicha información parcial, y calcular las correspondientes esperanzas condicionadas,

$$E(X \mid Y \leq y) , \qquad E(X \mid Y > y) , \qquad E(X \mid y_1 < Y \leq y_2) , \qquad \ldots\ldots ,$$

según los casos. En general, *si* B *es un conjunto de Borel de* β^2, *puede calcularse la esperanza de* X *condicionada por el suceso* B,

$$E(X \mid B) ,$$

como

$$E(X \mid B) = \sum_{i=1}^{\infty} x_i \cdot p[X = x_i \mid B] ,$$

si X es discreta, o bien

$$E(X \mid B) = \int_{x=-\infty}^{\infty} x \cdot f_{(X \mid B)}(x)\, dx\ ,$$

si es continua.

Ejemplo 7.4. Tenemos una urna con tres bolas blancas y dos negras. Se realizan tres extracciones sin reposición. Sea X la variable *número de bolas blancas extraídas*, e Y la variable *número de bolas negras antes de extraer la primera bola blanca*. Los valores que pueden tomar dichas variables son:

$$X: \left|\begin{array}{l}1\\2\\3\end{array}\right. \qquad \text{e} \qquad Y: \left|\begin{array}{l}0\\1\\2\end{array}\right.$$

Designemos por B_i y N_i a los sucesos *la bola extraída es blanca* y *la bola extraída es negra*, respectivamente, en la i-ésima extracción, para $i = 1, 2, 3$. Podemos calcular la función de probabilidad conjunta. Así, por ejemplo,

$$\begin{aligned} p[X=1, Y=0] = p(B_1 N_2 N_3) &= p(B_1)\cdot p(N_2 \mid B_1)\cdot p(N_3 \mid B_1 N_2) = \\ &= \frac{3}{5}\cdot\frac{2}{4}\cdot\frac{1}{3} = \frac{1}{10} \end{aligned}$$

y

$$\begin{aligned} p[X=1, Y=1] = p(N_1 B_2 N_3) &= p(N_1)\cdot p(B_2 \mid N_1)\cdot p(N_3 \mid N_1 B_2) = \\ &= \frac{2}{5}\cdot\frac{3}{4}\cdot\frac{1}{3} = \frac{1}{10}\ . \end{aligned}$$

Calculando del mismo modo el resto de las probabilidades, podemos completar la siguiente tabla de doble entrada.

$X \backslash Y$	0	1	2	
1	1/10	1/10	1/10	3/10
2	4/10	2/10	0	6/10
3	1/10	0	0	1/10
	6/10	3/10	1/10	1

El valor esperado de X, condicionado a que Y sea igual a cero, es decir, el número de bolas blancas extraídas cuando la primera lo es, viene dado por la siguiente expresión:

$$E(X \mid Y=0) = \sum_{i=1}^{3} x_i \cdot p[X = x_i \mid Y = 0]\ .$$

Obtengamos, previamente, la distribución condicionada,

$$p[X = x_i \mid Y = 0] = \frac{p[X = x_i, Y = 0]}{p[Y=0]} = \left\{\begin{array}{ll} 1/6 & x = 1 \\ 2/3 & x = 2 \\ 1/6 & x = 3 \end{array}\right.$$

con lo que la esperanza condicionada resultará:

$$E(X \mid Y=0) = 1\cdot\frac{1}{6} + 2\cdot\frac{2}{3} + 3\cdot\frac{1}{6} = 2 \text{ bolas blancas.}$$

Podemos, asimismo, calcular el valor esperado de Y cuando X es mayor que la unidad, es decir, el número esperado de bolas negras, antes de obtener la primera blanca, sabiendo que se ha obtenido más de una bola blanca. Hallemos antes la distribución de Y condicionada por el suceso $[X > 1]$,

$$\begin{aligned} f_{(Y \mid X>1)}(y_j) = p[Y = y_j \mid X > 1] &= \frac{p[X > 1, Y = y_j]}{p[X > 1]} = \\ &= \frac{p[X = 2, Y = y_j] + p[X = 3, Y = y_j]}{p[X = 2] + p[X = 3]} = \\ &= \begin{cases} 5/7 & y = 0 \\ 2/7 & y = 1 \\ 0 & y = 2 \end{cases} \end{aligned}$$

y, a partir de ella, la esperanza condicionada es

$$E(Y \mid X > 1) = \sum_{j=1}^{3} y_j \cdot p[Y = y_j \mid X > 1] = \frac{2}{7} \cdot \triangleleft$$

Ejemplo 7.5. Se está estudiando la relación entre la *renta disponible anual*, X, y el *consumo en bienes perecederos*, Y, en millones de pesetas, de las familias españolas. Se supone que su comportamiento viene explicado por la siguiente función de densidad conjunta:

$$f_{(X,Y)}(x, y) = \begin{cases} e^{-x} & 0 < y < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

¿Cuál es el consumo esperado de una familia que tiene una renta de 2.5 millones de pesetas? Nuestro objetivo es calcular

$$E(Y \mid X = 2.5) = \int_{y=-\infty}^{\infty} y \cdot f_{(Y \mid X=2.5)}(y)\, dy \,.$$

Obtengamos la distribución condicionada correspondiente,

$$f_{(Y \mid X=2.5)}(y) = \frac{f_{(X,Y)}(2.5, y)}{f_X(2.5)} \,.$$

La función de densidad de X es

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{(X,Y)}(x, y)dy = \int_{y=0}^{x} e^{-x}dy = \begin{cases} xe^{-x} & 0 < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

Por tanto,

$$f_{(Y \mid X=2.5)}(y) = \frac{e^{-2.5}}{2.5 \cdot e^{-2.5}} = \begin{cases} \frac{1}{2.5} & 0 < y < 2.5 \\ 0 & \text{en el resto} \end{cases}$$

La esperanza condicionada resulta:

$$E(Y \mid X = 2.5) = \int_{y=0}^{2.5} y \cdot \frac{1}{2.5} dy = 1.25 \text{ millones de pesetas.}$$

Evidentemente, si no fijamos un valor concreto de X, la esperanza condicionada será una función de dicho valor, x,

$$E(Y \mid X = x) = \int_{y=-\infty}^{\infty} y \cdot f_{(Y \mid X=x)}(y)\, dy = \int_{y=-\infty}^{\infty} y \cdot \frac{f_{(X,Y)}(x,y)}{f_X(x)} dy =$$
$$= \int_{y=0}^{x} y \cdot \frac{e^{-x}}{xe^{-x}} dy = \frac{x}{2} \qquad 0 < x < \infty .$$

Dando a x diferentes valores, calcularíamos el consumo esperado para distintas rentas disponibles de las familias.

Hallemos, ahora, el consumo esperado para las familias cuya renta no supera los dos millones, es decir,

$$E(Y \mid X \leq 2) = \int_{y=-\infty}^{\infty} y \cdot f_{(Y \mid X \leq 2)}(y)\, dy .$$

Como vimos en el capítulo 5, para hallar la función de densidad condicionada hemos de calcular primero la función de distribución correspondiente,

$$F_{(Y \mid X \leq 2)}(y) = p[Y \leq y \mid X \leq 2] = \frac{p[X \leq 2, Y \leq y]}{p[X \leq 2]} ,$$

donde

$$p[X \leq 2] = \int_{x=0}^{2} f_X(x)dx = \int_{x=0}^{2} xe^{-x}dx = 1 - 3e^{-2} = 0.59$$

y, si $0 < y < 2$,

$$p[X \leq 2, Y \leq y] = \int_{t=0}^{y}\int_{s=t}^{2} f_{(X,Y)}(s,t)dsdt = \int_{t=0}^{y}\int_{s=t}^{2} e^{-s}dsdt =$$
$$= 1 - e^{-y} - e^{-2}y .$$

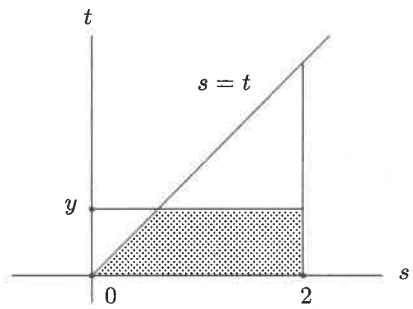


FIG. 7.3.

Por tanto,

$$F_{(Y \mid X \leq 2)}(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1-e^{-y}-e^{-2}y}{1-3e^{-2}} & 0 < y < 2 \\ 1 & y \geq 2 \end{cases}$$

Derivando, obtenemos la función de densidad condicionada,

$$f_{(Y \mid X \leq 2)}(y) = \begin{cases} \frac{e^{-y}-e^{-2}}{1-3e^{-2}} & 0 < y < 2 \\ 0 & \text{en el resto} \end{cases}$$

En consecuencia,

$$E(Y \mid X \leq 2) = \int_{y=0}^{2} y \cdot \frac{e^{-y} - e^{-2}}{1 - 3e^{-2}} dy = \frac{1 - 5e^{-2}}{1 - 3e^{-2}} = 0.54 \text{ millones de pesetas.} \triangleleft$$

Veamos, a continuación, algunas propiedades de la esperanza condicionada.

P1. Para un valor concreto de Y, y, la esperanza de X condicionada por $Y = y$, $E(X \mid Y = y)$, es la esperanza de una distribución, verificando, por tanto, todas las propiedades generales de la esperanza de una variable unidimensional que estudiamos en el capítulo anterior. Citaremos algunas de ellas, sin ánimo de ser exhaustivos.

a) Si X es una variable degenerada, es decir, una constante, su esperanza condicionada por $Y = y$ es ella misma; esto es, si $p[X = K] = 1$,

$$E(X \mid Y = y) = K \, .$$

b) La esperanza condicionada verifica la propiedad de linealidad. Es decir,

$$E[(a \cdot g(X) + b \cdot h(X)) \mid Y = y] = a \cdot E[g(X) \mid Y = y] + b \cdot E[h(X) \mid Y = y] \, ,$$

siempre que existan dichas esperanzas.

c) Si $g(X) \leq h(X)$, entonces $E[g(X) \mid Y = y] \leq E[h(X) \mid Y = y]$, siempre que existan dichas esperanzas.

Las dos propiedades siguientes son nuevas, en el sentido de que provienen no del hecho de que la esperanza condicionada es una esperanza, sino de la naturaleza condicionada de la misma.

P2. Si X e Y son variables aleatorias independientes,

$$E(X \mid Y = y) = E(X) \, ,$$

siempre que dichas esperanzas existan.

Demostración. Ya vimos en el capítulo 5 que si X e Y son independientes,

$$f_{(X \mid Y=y)}(x) = f_X(x) \, .$$

Entonces,

$$E(X \mid Y = y) = \int_{x=-\infty}^{\infty} x \cdot f_{(X \mid Y=y)}(x) dx = \int_{x=-\infty}^{\infty} x \cdot f_X(x) dx = E(X) \, .$$

El mismo razonamiento emplearíamos si X e Y fueran variables discretas. ♣

P3. Sea (X,Y) una variable aleatoria bidimensional, y $h(X)$ una función de X que es, a su vez, variable aleatoria. Entonces,

$$E[E\,(h(X) \mid Y = y)] = E[h(X)]\,.$$

Demostración. Hallemos $E[E\,(h(X) \mid Y = y)]$, teniendo en cuenta que en esta expresión hay dos esperanzas. La interior es la esperanza condicionada que hemos definido en esta sección, es decir, la esperanza de $h(X)$ cuando X tiene la distribución condicionada por $Y = y$. El resultado es una función de y, cuando y recorre los valores de la variable aleatoria Y. Podemos calcular, entonces, la esperanza de dicha función de y, pero entendiendo que dicha función lo es de la variable aleatoria Y (de hecho, en estas situaciones, a veces se utiliza la notación $E[E(h(X) \mid Y)]$). En definitiva, la esperanza exterior se calcula con la distribución de probabilidades de Y. Por tanto,

$$\begin{aligned}
E[E(h(X) \mid Y = y)] &= \int_{y=-\infty}^{\infty} E[h(X) \mid Y = y] \cdot f_Y(y)dy = \\
&= \int_{y=-\infty}^{\infty} \left[\int_{x=-\infty}^{\infty} h(x) \cdot f_{(X \mid Y=y)}(x)dx\right] \cdot f_Y(y)dy = \\
&= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} h(x) \cdot f_{(X,Y)}(x,y)\; dydx = \\
&= \int_{x=-\infty}^{\infty} h(x) \cdot f_X(x)dx = E[h(X)]\,.
\end{aligned}$$

De igual manera se realizaría la demostración para variables discretas. ♣

Teniendo en cuenta la observación 4 de esta misma sección, y la propiedad anterior, es inmediata la siguiente igualdad:

$$E[E(h(X,Y) \mid X = x)] = E[h(X,Y)]\,.$$

Asimismo, de la propiedad **3** se deduce que, si $h(X) = X$, entonces,

$$E[E(X \mid Y = y)] = E(X)\,.$$

La propiedad anterior permite demostrar un resultado que es, parcialmente, una propiedad recíproca de la segunda.

P4. Si la esperanza de X condicionada por $Y = y$ es constante, es decir, $E(X \mid Y = y) = K$, las variables X e Y están incorrelacionadas.

Demostración. Si $E(X \mid Y = y) = K$, siendo K una constante cualquiera,

$$E[E(X \mid Y = y)] = E(K) = K\,.$$

Pero la propiedad anterior indica que dicha esperanza doble es la esperanza de X, y, consecuentemente, se cumple que $E(X) = K$. En definitiva,

$$E(X \mid Y = y) = E(X).$$

Comprobemos que, entonces, la covarianza se anula.

$$\begin{aligned} E(X\cdot Y) &= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} x\cdot y\cdot f_{(X,Y)}(x,y)\,dydx = \\ &= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} x\cdot y\cdot f_{(X\mid Y=y)}(x)\cdot f_Y(y)\,dydx = \\ &= \int_{y=-\infty}^{\infty} y\cdot f_Y(y)\left[\int_{x=-\infty}^{\infty} x\cdot f_{(X\mid Y=y)}(x)\,dx\right]dy = \\ &= \int_{y=-\infty}^{\infty} y\cdot f_Y(y)[E(X\mid Y=y)]dy = E(X)\cdot\int_{y=-\infty}^{\infty} y\cdot f_Y(y)dy = \\ &= E(X)\cdot E(Y)\,. \end{aligned}$$

De este modo, $E(X\cdot Y)=E(X)\cdot E(Y)$ y, por consiguiente, la covarianza de X e Y es igual a cero. Las variables están, pues, incorrelacionadas.

En el caso de que X e Y fuesen discretas el razonamiento sería análogo. ♣

Las anteriores definiciones de esperanza condicionada se pueden generalizar a n variables. Es decir, *dada una variable aleatoria n-dimensional $(X_1,\ldots,X_n)$, definimos la esperanza de una de sus componentes, por ejemplo X_1, condicionada a ciertos valores del resto de las variables, $E(X_1\mid X_2=x_2,\ldots,X_n=x_n)$, como el valor*

$$\sum_{i=1}^{\infty} x_{1i}\cdot p[X_1=x_{1i}\mid X_2=x_2,\ldots,X_n=x_n]\,,$$

si $(X_1,\ldots,X_n)$ es una variable discreta, y

$$\int_{x_1=-\infty}^{\infty} x_1\cdot f_{(X_1\mid X_2=x_2,\ldots,X_n=x_n)}(x_1)\,dx_1\,,$$

si $(X_1,\ldots,X_n)$ es continua. Dicha definición está sujeta a los ya citados problemas de existencia.

7.3 Momentos condicionados

En el capítulo 6 definimos los momentos de una distribución como medidas que resumen ciertos aspectos de la misma. Vamos a definir, ahora, los momentos de las distribuciones condicionadas, concretamente los momentos centrados respecto del origen y respecto de la media. Como comentamos al estudiar la esperanza condicionada, estas definiciones no son nuevas, porque los momentos condicionados son los momentos de una variable unidimensional cuando su distribución está condicionada.

Definición 7.7 *Sea (X,Y) una variable aleatoria bidimensional. Llamaremos momento de X condicionado por $Y=y$ centrado respecto del origen de orden k $(k=0,1,\ldots)$, o momento condicionado no central, a la esperanza, si existe,*

$$E(X^k \mid Y = y) = \sum_{i=1}^{\infty} x_i^k \cdot p[X = x_i \mid Y = y] ,$$

si la distribución condicionada es discreta, o bien

$$E(X^k \mid Y = y) = \int_{x=-\infty}^{\infty} x^k \cdot f_{(X \mid Y=y)}(x)dx ,$$

si la distribución condicionada es continua.

Si $k = 0$,

$$E(X^0 \mid Y = y) = E(1 \mid Y = y) = 1 ,$$

por lo que dicho momento no informa sobre la distribución condicionada al tomar siempre el mismo valor.

Para $k = 1$,

$$E(X \mid Y = y) ,$$

es decir, se obtiene la esperanza de X condicionada al valor y de Y.

Para $k = 2$,

$$E(X^2 \mid Y = y) ,$$

que se utiliza en el cálculo de la varianza condicionada que definiremos más adelante.

Definición 7.8 *Sea (X, Y) una variable aleatoria bidimensional. Llamaremos momento de X condicionado por $Y = y$ centrado respecto de la media condicionada de orden k ($k = 0, 1, \ldots$), o momento condicionado central, a la esperanza, si existe,*

$$E[(X - E(X \mid Y = y))^k \mid Y = y] ,$$

que, si la distribución condicionada es discreta, será igual a

$$\sum_{i=1}^{\infty} [x_i - E(X \mid Y = y)]^k \cdot p[X = x_i \mid Y = y]$$

y, si la distribución condicionada es continua, a

$$\int_{x=-\infty}^{\infty} [x - E(X \mid Y = y)]^k \cdot f_{(X \mid Y=y)}(x)\, dx .$$

Si $k = 0$,

$$E[(X - E(X \mid Y = y))^0 \mid Y = y] = E[1 \mid Y = y] = 1 ,$$

que, al igual que el correspondiente momento respecto del origen, no informa sobre la distribución condicionada.

Para $k = 1$,

$$\begin{aligned} E[(X - E(X|Y = y)) \mid Y = y] &= E(X \mid Y = y) - E[E(X|Y = y) \mid Y = y] = \\ &= E(X \mid Y = y) - E(X \mid Y = y) = 0, \end{aligned}$$

ya que $E(X \mid Y = y)$ es una constante para un valor concreto de y, y por tanto, $E[E(X \mid Y = y) \mid Y = y] = E(X \mid Y = y)$.

Este momento condicionado tampoco aporta información sobre ningún aspecto de la distribución condicionada.

Para $k = 2$, se obtiene el momento denominado **varianza de X condicionada por $Y = y$**,

$$E[(X - E(X \mid Y = y))^2 \mid Y = y] = Var(X \mid Y = y) .$$

Dicho momento, función de y, mide la dispersión de la distribución de X condicionada por $Y = y$ respecto de su media condicionada, $E(X \mid Y = y)$.

Como hicimos en el capítulo anterior, podríamos definir otro tipo de momentos condicionados (respecto de C, absolutos y potenciales). Su interpretación es paralela a la de aquellos momentos, refiriéndose, en este caso, a las distribuciones condicionadas.

Estudiemos, seguidamente, las propiedades del momento condicionado central más importante, la varianza condicionada.

P1. Dada una variable bidimensional (X, Y), la varianza de X condicionada por $Y = y$ se puede descomponer como

$$Var(X \mid Y = y) = E(X^2 \mid Y = y) - [E(X \mid Y = y)]^2 .$$

Demostración. Dicho resultado es inmediato ya que, para una distribución cualquiera,

$$Var(X) = E(X^2) - [E(X)]^2 ,$$

siendo la propiedad **1**, una concreción del resultado anterior para una distribución condicionada. Aun así, reiteremos la demostración.

$$Var(X \mid Y = y) = E[(X - E(X \mid Y = y))^2 \mid Y = y] ,$$

que, desarrollando el cuadrado, resulta

$$E(X^2 \mid Y = y) + E[(E(X \mid Y = y))^2 \mid Y = y] - 2E(X \mid Y = y)E(X \mid Y = y) .$$

Entonces,

$$\begin{aligned} Var(X \mid Y = y) &= E(X^2 \mid Y = y) + [E(X \mid Y = y)]^2 - 2[E(X \mid Y = y)]^2 = \\ &= E(X^2 \mid Y = y) - [E(X \mid Y = y)]^2 . \clubsuit \end{aligned}$$

Como ocurría para la varianza de una variable, la anterior descomposición de la varianza de X condicionada por $Y = y$ facilitará su cálculo.

La varianza de una variable se puede descomponer en función de la varianza condicionada. Más concretamente, puede expresarse como suma de la esperanza de la varianza condicionada y de la varianza de la esperanza condicionada.

P2. Sea (X,Y) una variable aleatoria bidimensional. La varianza de X se puede descomponer de la siguiente forma:

$$Var(X) = E\left[Var(X \mid Y=y)\right] + Var\left[E(X \mid Y=y)\right] .$$

Demostración. Desarrollando los dos sumandos,

$$\begin{aligned} E\left[Var(X \mid Y=y)\right] &= E\left[E(X^2 \mid Y=y) - (E(X \mid Y=y))^2\right] = \\ &= E(X^2) - E\left[(E(X \mid Y=y))^2\right] \end{aligned}$$

y

$$\begin{aligned} Var\left[E(X \mid Y=y)\right] &= E[(E(X \mid Y=y))^2] - [E\left(E(X \mid Y=y)\right)]^2 = \\ &= E\left[(E(X \mid Y=y))^2\right] - [E(X)]^2 . \end{aligned}$$

Sumando miembro a miembro ambas descomposiciones obtenemos:

$$E\left[Var(X \mid Y=y)\right] + Var\left[E(X \mid Y=y)\right] = E(X^2) - [E(X)]^2 = Var(X) . \clubsuit$$

Ejemplo 7.6. Consideremos, de nuevo, la función de densidad conjunta del ejemplo 7.5:

$$f_{(X,Y)}(x,y) = \begin{cases} e^{-x} & 0 < y < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

Hallemos el valor de

$$Var(Y \mid X=3) .$$

Aplicando la propiedad **1**,

$$Var(Y \mid X=3) = E(Y^2 \mid X=3) - [E(Y \mid X=3)]^2 .$$

Calculemos cada uno de los sumandos, obteniendo previamente la distribución de Y condicionada por $X=3$. Teniendo en cuenta la función de densidad marginal de X, $f_X(x)$, hallada en el ejercicio 7.5,

$$f_{(Y \mid X=3)}(y) = \frac{f_{(X,Y)}(3,y)}{f_X(3)} = \frac{e^{-3}}{3 \cdot e^{-3}} = \frac{1}{3} , \qquad \text{si } 0 < y < 3 .$$

Así, la esperanza condicionada vale:

$$E(Y \mid X=3) = \int_{-\infty}^{\infty} y \cdot f_{(Y \mid X=3)}(y)dy = \int_{y=0}^{3} y \cdot \frac{1}{3} dy = \frac{3}{2} .$$

Por lo que respecta a $E(Y^2 \mid X=3)$, momento de segundo orden respecto al origen,

$$E(Y^2 \mid X=3) = \int_{y=-\infty}^{\infty} y^2 \cdot f_{(Y \mid X=3)}(y)\, dy = \int_{y=0}^{3} y^2 \cdot \frac{1}{3} dy = 3 .$$

Deducimos, así, que

$$Var(Y \mid X=3) = 3 - \left(\frac{3}{2}\right)^2 = 0.75 ,$$

valor que mide la dispersión de la variable *consumo en bienes perecederos* entre las familias con una renta disponible de 3 millones de pesetas. ◁

Al igual que ocurría con la esperanza condicionada, los momentos condicionados pueden generalizarse a n variables.

7.4 Regresión y correlación

Como señalábamos en la introducción de este capítulo, con frecuencia estamos interesados en el estudio de la relación existente entre dos (o más) variables aleatorias, de cara a explicar una de ellas a partir de la otra (u otras) y, en última instancia, poder predecir el valor de una cuando conocemos el que toma la otra (o el resto). Así, nos puede interesar estudiar la relación que existe entre el peso y la altura de un individuo, entre la renta y el consumo de una familia o entre el tamaño de una empresa y su inversión en investigación y desarrollo.

Algunas relaciones entre variables (principalmente en el campo de las ciencias físicas o químicas) son deterministas o se consideran como tales, pudiéndose describir de forma exacta mediante una expresión matemática. De esta forma, podemos hallar la longitud de una barra metálica en función de la temperatura a la que está sometida, o el volumen que ocupa un gas bajo unas ciertas condiciones de presión y temperatura.

Ahora bien, en muchas ocasiones es imposible aceptar la naturaleza determinista de las relaciones entre las variables, por lo que resulta más práctica la búsqueda de relaciones matemáticas aproximadas entre las mismas, entendiendo esta aproximación en términos probabilísticos. Así, la mayoría de las relaciones que se estudian en las ciencias sociales no admiten una reducción a simples expresiones matemáticas exactas. El consumo de una familia depende, evidentemente, de su renta, pero no podemos deducir de este hecho una *fórmula* que concrete el consumo de una familia con un determinado nivel de renta. En estas condiciones, trataremos de encontrar relaciones matemáticas que expliquen, de la mejor manera posible, una variable a partir de otra (u otras). Las técnicas que permiten hacerlo se agrupan bajo el nombre genérico de **regresión.**

El término *regresión* se acuñó a finales del siglo pasado por los investigadores en genética. Se denomina *regresión a los valores medios* al fenómeno por el cual, estadísticamente, la intensidad de un determinado carácter en un individuo es más próxima al valor medio de la población de lo que era en sus progenitores.

El origen del término se debe a Sir F. GALTON. Este biólogo, primo del naturalista CH. DARWIN, es uno de los principales impulsores de las aplicaciones estadísticas en el campo de la biología y de la genética. Ya en 1876, estudiando la efectividad de un tratamiento químico para las plantas a partir de datos recogidos por DARWIN, propuso un test de rangos que hoy día, con el nombre de test de las rachas, se utiliza en otros muchos campos de investigación.

Se conocen, asimismo, sus estudios en 1889 referentes a la supervivencia de los apellidos (aplicados, sesenta años más adelante, al análisis estadístico de las reacciones nucleares en cadena por E. SCHROEDINGER), que dieron origen a los denominados procesos de ramificación.

No obstante, dentro de la Estadística, es más conocido por sus estudios biométricos acerca de cómo se traslada la herencia genética de padres a hijos; estos estudios fueron realizados con K. PEARSON (a quien se debe el término *biometría*). Comparando la estatura de los padres con la de los hijos, apreció una tendencia hacia la estatura media, esto es, que los padres excepcionalmente bajos o excepcionalmente altos tienen normalmente hijos cuyas alturas se acercan al promedio.

GALTON sintetizó esta apreciación diciendo que las estaturas *regresan* o revierten hacia la media (*regression towards mediocrity in hereditary stature*)[2]. A partir de entonces, la palabra *regresión* se utilizó, impropiamente, para designar las técnicas de análisis de las relaciones entre variables estadísticas y, por extensión, las relaciones entre variables aleatorias.

En el campo económico, como decíamos en la introducción, los análisis relacionados con la regresión provocaron, a comienzos de los años treinta, el nacimiento de una rama científica que, por la peculiaridad de las técnicas que ha diseñado y de su campo de aplicación, se ha desgajado parcialmente de la Estadística; hablamos de la Econometría, disciplina que constituye hoy en día una herramienta básica en la investigación económica.

Todos los desarrollos de la regresión los realizaremos para dos variables; el lector podrá intuir el sentido y alcance de la posible generalización a n variables cuando se trata de explicar una variable en función del resto.

Dada una variable aleatoria bidimensional (X, Y), buscaremos una función de X que explique todo cuanto pueda de Y o una función de Y que explique lo mejor posible X. Por concretar, supongamos que estamos en la primera de estas dos situaciones. Sea $g(X)$ una función que proporciona algo *parecido* a Y; a esta función de X la llamaremos Y *teórico*[3], designándolo por $\tilde{Y}$:

$$\tilde{Y} = g(X) \, .$$

Al considerar $\tilde{Y}$ como Y cometemos un error ε,

$$\varepsilon = Y - \tilde{Y} = Y - g(X) \, .$$

Téngase en cuenta que este error ε no es un valor concreto, sino que es una variable aleatoria función de X y de Y. En la búsqueda de esa función de X que proporcione una buena explicación de Y, trataremos de que el error que cometamos sea lo más pequeño posible, es decir, que $\tilde{Y}$ sea lo más parecido a Y. Intentaremos, por tanto, buscar la función $g(X)$ que haga el error *lo más pequeño posible*. Ahora bien, advierta el lector que, como decíamos, este error es una variable aleatoria, que toma distintos valores con ciertas probabilidades. ¿En qué sentido podemos hablar de un error pequeño? Existen diversos criterios para *medir* este error y, por consiguiente, poder encontrar una función $g(X)$ que lo minimice. El criterio más usual es el **criterio de los mínimos cuadrados**. Mediremos el error en términos de la esperanza de su cuadrado; es decir, cuantificaremos el error en media,

2. GALTON, F.: *Natural Inheritance*, Londres, Macmillan, 1889.

3. Denominar Y *teórico* a la función $g(X)$ que se asemeja a Y es una costumbre cuyos orígenes se remontan a los análisis descriptivos de regresión. En éstos, se dispone de una nube de puntos (x_i, y_j) que representan los valores *reales* que toman dos variables, X e Y, para una población *real*. Si $g(X)$ es la función de X que mejor explica Y, a sus valores, $\tilde{y}_i = g(x_i)$, se les llama con propiedad valores *teóricos* de Y, porque son los valores que tomaría la variable Y para la población en estudio si la relación $Y = g(X)$ fuese cierta.

Ahora bien, las variables aleatorias que estudiamos en este libro son siempre modelos teóricos de lo que ocurre en los experimentos aleatorios, por lo que podría ser merecedora de ese calificativo tanto la variable Y como la variable $\tilde{Y} = g(X)$.

considerando los valores de ε elevados al cuadrado para evitar que se compensen los errores cometidos por exceso y los cometidos por defecto. A continuación, buscaremos la función g que minimice dicha cantidad, esto es, una función g de la variable aleatoria X, tal que

$$\min_g E(\varepsilon^2) = \min_g E[(Y - g(X))^2] .$$

Una vez elegido el criterio, podemos enfocar el problema de la búsqueda de la función g de dos maneras:

1) *Regresión sin restricciones*

Buscaremos g entre todas las posibles funciones, sin someterla a ningún tipo de restricciones. La función obtenida de esta forma recibe el nombre de **función** o **curva general de regresión de** Y **sobre** X y proporciona la mejor explicación posible de Y a partir de X. Esta función no tiene por qué ser una función de uso común, lo que puede dificultar su interpretación.

2) *Regresión con restricciones*

En este caso, buscaremos g dentro de un cierto tipo de funciones, fundamentalmente entre aquellas más estudiadas, conocidas y manejables. Por ejemplo:

- Si exigimos que $g(X)$ sea una función lineal, es decir, $g(X) = \alpha \cdot X + \beta$, encontraremos la recta que mejor explica Y a partir de X: **la recta de regresión de** Y **sobre** X.
- Si exigimos que $g(X)$ sea una función parabólica, $g(X) = \alpha \cdot X^2 + \beta \cdot X + \gamma$, encontraremos la parábola de segundo grado que mejor explica Y a partir de X: **la parábola de regresión de** Y **sobre** X.

Igualmente podríamos hacer restricciones de otro tipo.

Esta forma de plantearse la búsqueda de $g(X)$ conduce a funciones que no explican tan bien la variable Y a partir de X como lo hace la función general de regresión. Pero, a su favor, presenta la ventaja de proporcionar funciones de una interpretación más sencilla y evidente.

Siguiendo uno u otro enfoque, conseguiremos la función, la recta o la parábola que mejor explica Y a partir de X. Aun así, dicha explicación puede no ser buena. Nos plantearemos, en consecuencia, estudiar la bondad de la explicación de Y proporcionada por $g(X)$, tras haber seleccionado adecuadamente la función g.

Con el fin de simplificar, utilizaremos la misma notación, $\tilde{Y} = g(X)$, para designar tanto a una función genérica que explica Y a partir de X, como a la función concreta que proporciona la mejor explicación; mantendremos el mismo criterio para denotar los errores de la regresión.

Para facilitar la comprensión del lector, comenzaremos el estudio de la regresión poniendo restricciones a la forma funcional de $g(X)$. En concreto, desarrollaremos el caso de mayor interés, exigiendo que $g(X)$ sea una función lineal, una recta.

Para cualquier otro tipo de restricción (que no plantearemos en este libro), los desarrollos son totalmente paralelos a los aquí presentados.

Para estudiar los problemas de regresión que se plantearán en las dos siguientes subsecciones, supondremos que tanto la varianza de la variable X como la de Y son distintas de cero. Esta suposición es necesaria desde el punto de vista analítico, pero, además, es razonable desde el punto de vista intuitivo. Como hemos indicado anteriormente, la regresión trata de obtener funciones que resuman la relación entre dos variables aleatorias; más en concreto, su objetivo es explicar la variabilidad de una variable en función de la de la otra. Pero, si la primera es constante, no hay nada que explicar, y si lo es la segunda, la ausencia de variaciones de ésta impide estudiar las de la primera.

7.4.1 Regresión lineal

Sea una variable bidimensional (X, Y). Trataremos de encontrar la función lineal de X, $\alpha \cdot X + \beta$, que mejor explique Y, de acuerdo con el criterio mínimo cuadrático. Sea $\tilde{Y}$ una explicación lineal de Y,

$$\tilde{Y} = \alpha \cdot X + \beta \,,$$

y sea ε, la variable aleatoria que representa el error cometido al tomar $\tilde{Y}$ por el verdadero Y,

$$\varepsilon = Y - \tilde{Y} = Y - \alpha \cdot X - \beta \,.$$

Busquemos los valores de α y β, parámetros que determinan la función lineal, con la condición de que minimicen la esperanza de la variable ε al cuadrado:

$$\min_{\alpha,\beta} E(\varepsilon^2) = \min_{\alpha,\beta} E[(Y - \alpha \cdot X - \beta)^2] \,.$$

El problema se concreta en minimizar una función que depende de dos variables, α y β:

$$\Phi(\alpha, \beta) = E[(Y - \alpha \cdot X - \beta)^2] \,.$$

Desarrollando el cuadrado que aparece en la expresión anterior y teniendo en cuenta la linealidad de la esperanza,

$$\begin{aligned} \Phi(\alpha,\beta) &= E[(Y - \alpha \cdot X - \beta)^2] = \\ &= E(Y^2 - 2\alpha X \cdot Y - 2\beta Y + \alpha^2 X^2 + 2\alpha\beta X + \beta^2) = \\ &= E(Y^2) - 2\alpha E(X \cdot Y) - 2\beta E(Y) + \alpha^2 E(X^2) + 2\alpha\beta E(X) + \beta^2 \,. \end{aligned}$$

Entonces, derivando Φ respecto de α y de β e igualando a cero, obtenemos:

$$\frac{\partial \Phi(\alpha,\beta)}{\partial \alpha} = -2E(X \cdot Y) + 2\alpha E(X^2) + 2\beta E(X) = 0$$

y

$$\frac{\partial \Phi(\alpha,\beta)}{\partial \beta} = -2E(Y) + 2\alpha E(X) + 2\beta = 0 \,.$$

Dividiendo por -2 ambas ecuaciones, resulta

$$E(X \cdot Y) - \alpha \cdot E(X^2) - \beta \cdot E(X) = 0$$

y

$$E(Y) - \alpha \cdot E(X) - \beta = 0 .$$

Por último, multiplicando la segunda ecuación por $E(X)$ y restando el resultado a la primera, obtenemos:

$$E(X \cdot Y) - E(X) \cdot E(Y) - \alpha \cdot [E(X^2) - (E(X))^2] = 0 ,$$

es decir,

$$Cov(X, Y) - \alpha \cdot Var(X) = 0 .$$

De aquí, el valor de α que minimiza la $E(\varepsilon^2)$ será

$$\alpha = \frac{Cov(X, Y)}{Var(X)} .$$

Sustituyendo en la segunda de las ecuaciones obtenemos el valor correspondiente de β:

$$\beta = E(Y) - \frac{Cov(X, Y)}{Var(X)} \cdot E(X) .$$

Comprobemos que en estos valores α y β se alcanza realmente el mínimo, para lo cual calculemos el hessiano correspondiente:

$$H = \begin{vmatrix} \dfrac{\partial^2 \Phi}{\partial \alpha^2} & \dfrac{\partial^2 \Phi}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 \Phi}{\partial \beta \partial \alpha} & \dfrac{\partial^2 \Phi}{\partial \beta^2} \end{vmatrix} ,$$

donde

$$\frac{\partial^2 \Phi}{\partial \alpha^2} = 2E(X^2) ,$$
$$\frac{\partial^2 \Phi}{\partial \beta^2} = 2 \quad \text{y}$$
$$\frac{\partial^2 \Phi}{\partial \beta \partial \alpha} = \frac{\partial^2 \Phi}{\partial \alpha \partial \beta} = 2E(X) .$$

Por tanto,

$$H = \begin{vmatrix} 2E(X^2) & 2E(X) \\ 2E(X) & 2 \end{vmatrix} .$$

Estudiando el signo de los menores principales del determinante, se comprueba que ambos son estrictamente positivos, lo que confirma que dichos valores α y β son los que minimizan el error cuadrático medio.

Entonces, la función lineal de X, $\tilde{Y} = \alpha \cdot X + \beta$, que minimiza $E(\varepsilon^2)$ es:

$$\tilde{Y} = \alpha \cdot X + \beta = \frac{Cov(X, Y)}{Var(X)} \cdot X + E(Y) - \frac{Cov(X, Y)}{Var(X)} \cdot E(X) .$$

Dicha expresión, que suele presentarse de la siguiente forma:

$$\tilde{Y} - E(Y) = \frac{Cov(X,Y)}{Var(X)} \cdot [X - E(X)] ,$$

recibe el nombre de **recta de regresión mínimo cuadrática de Y sobre X.**

El coeficiente de X en dicha función se denomina coeficiente de regresión de Y sobre X:

$$b_{Y \mid X} = \frac{Cov(X,Y)}{Var(X)} .$$

En muchas ocasiones, nos referiremos a la recta de regresión no como relación lineal entre variables aleatorias, sino como la expresión matemática de una recta. En este caso, utilizaremos la notación habitual de las funciones matemáticas:

$$y - E(Y) = \frac{Cov(X,Y)}{Var(X)} \cdot [x - E(X)] .$$

El coeficiente de regresión, $b_{Y \mid X}$, no es otra cosa que la pendiente de dicha recta, que designaremos como $p_{Y \mid X}$. Obviamente,

$$p_{Y \mid X} = \frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_X} \cdot \frac{\sigma_Y}{\sigma_Y} = \rho \cdot \frac{\sigma_Y}{\sigma_X} ,$$

con lo cual podemos expresar la recta de regresión en función del coeficiente de correlación:

$$y - \mu_Y = \rho \cdot \frac{\sigma_Y}{\sigma_X} \cdot (x - \mu_X) .$$

Haciendo el mismo planteamiento para explicar X a partir de una función lineal de Y, $\tilde{X} = \alpha \cdot Y + \beta$, obtendríamos:

$$\tilde{X} - E(X) = \frac{Cov(X,Y)}{Var(Y)} \cdot [Y - E(Y)] ,$$

que será la **recta de regresión de X sobre Y**, con

$$b_{X \mid Y} = \frac{Cov(X,Y)}{Var(Y)}$$

el correspondiente coeficiente de regresión.

Si la expresamos como función entre variables matemáticas, es decir, como una recta, tendremos

$$x - E(X) = \frac{Cov(X,Y)}{Var(Y)} \cdot [y - E(Y)] .$$

En este caso, la pendiente de la recta de regresión, $p_{X \mid Y}$, no es el coeficiente de regresión, sino su valor inverso,

$$p_{X \mid Y} = \frac{1}{b_{X \mid Y}} ,$$

que existe si $Cov(X,Y) \neq 0$. Asimismo, podemos expresar esta recta como

$$x - \mu_X = \rho \cdot \frac{\sigma_X}{\sigma_Y} \cdot (y - \mu_Y) .$$

Se comprueba fácilmente que ambas rectas pasan por el centro de gravedad de la nube de puntos, esto es, $(E(X), E(Y))$. Por tanto, o bien las dos rectas de regresión se cortan en dicho punto, o bien coinciden. La figura 7.4 ilustra este hecho sobre la nube de puntos de una variable aleatoria bidimensional discreta.

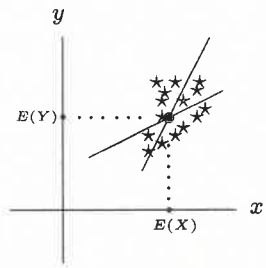


FIG. 7.4

Cuando realizamos la regresión de Y sobre X, partimos de dos variables, X e Y, de las cuales conocemos, al menos, sus esperanzas, varianzas y covarianza. En el proceso de regresión se generan dos nuevas variables: $\tilde{Y}$, es decir, Y *teórico*, y ε, error de la regresión. Estudiemos las principales características de estas dos variables. Así, puesto que $\tilde{Y}$ es una función lineal de X,

$$\tilde{Y} = \frac{Cov(X,Y)}{Var(X)} \cdot X + \left[E(Y) - \frac{Cov(X,Y)}{Var(X)} \cdot E(X) \right] ,$$

su esperanza será:

$$E(\tilde{Y}) = \frac{Cov(X,Y)}{Var(X)} \cdot E(X) + \left[E(Y) - \frac{Cov(X,Y)}{Var(X)} \cdot E(X) \right] = E(Y) .$$

Lo obtenido confirma que si queremos explicar Y con $\tilde{Y}$, parece razonable que, al menos, *acertemos* en media. Por otro lado, y considerando de nuevo $\tilde{Y}$ como función lineal de X, se tiene:

$$Var(\tilde{Y}) = \left[\frac{Cov(X,Y)}{Var(X)} \right]^2 \cdot Var(X) = \frac{[Cov(X,Y)]^2}{Var(X)} .$$

En cuanto a la variable ε,

$$\varepsilon = Y - \tilde{Y} = [Y - E(Y)] - \frac{Cov(X,Y)}{Var(X)} \cdot [X - E(X)] ,$$

su esperanza será:

$$E(\varepsilon) = E(Y - \tilde{Y}) = E(Y) - E(\tilde{Y}) = 0 \,,$$

resultado igualmente razonable, pues parece lógico que no haya errores sistemáticos, de forma que los que se producen por exceso (errores negativos) quedarán compensados por los que lo hacen por defecto (errores positivos). En cuanto a la varianza de los errores,

$$\begin{aligned} Var(\varepsilon) &= E(\varepsilon^2) = E\left[\left((Y - E(Y)) - \tfrac{Cov(X,Y)}{Var(X)} \cdot (X - E(X))\right)^2\right] = \\ &= E[(Y - E(Y))^2] + \left[\tfrac{Cov(X,Y)}{Var(X)}\right]^2 \cdot E[(X - E(X))^2] - \\ &\quad - 2\tfrac{Cov(X,Y)}{Var(X)} E[(X - E(X)) \cdot (Y - E(Y))] = \\ &= Var(Y) + \left[\tfrac{Cov(X,Y)}{Var(X)}\right]^2 \cdot Var(X) - 2\tfrac{Cov(X,Y)}{Var(X)} Cov(X,Y) = \\ &= Var(Y) - \tfrac{[Cov(X,Y)]^2}{Var(X)} \,. \end{aligned}$$

Este resultado se concreta de dos maneras. En primer lugar, y teniendo en cuenta que, como hemos demostrado,

$$Var(\tilde{Y}) = \frac{[Cov(X,Y)]^2}{Var(X)} \,,$$

entonces,

$$Var(\varepsilon) = Var(Y) - Var(\tilde{Y}) \,,$$

o, lo que es lo mismo,

$$Var(Y) = Var(\tilde{Y}) + Var(\varepsilon) \,,$$

expresión conocida como **descomposición de la varianza**, y que proporciona la varianza de la variable a explicar, $Var(Y)$, como suma de la varianza explicada por la regresión, $Var(\tilde{Y})$, y la varianza no explicada o residual, $Var(\varepsilon)$.

En segundo lugar, sacando factor común al término $Var(Y)$,

$$\begin{aligned} Var(\varepsilon) &= Var(Y) - \tfrac{[Cov(X,Y)]^2}{Var(X)} = Var(Y) \cdot \left[1 - \tfrac{(Cov(X,Y))^2}{Var(X)Var(Y)}\right] = \\ &= Var(Y) \cdot (1 - \rho^2) \,. \end{aligned}$$

El lector puede deducir de esta expresión, y de forma más sencilla que en el apartado 7.1, que el coeficiente de correlación lineal, ρ, está comprendido entre -1 y 1.

Hemos hallado la recta de regresión de Y sobre X bajo la condición de ser la recta que minimice la esperanza del cuadrado del error. Dado que $E(\varepsilon) = 0$, ese mínimo es precisamente

$$E(\varepsilon^2) = Var(Y) - Var(\tilde{Y}) = Var(Y) \cdot (1 - \rho^2) \,,$$

es decir, cualquier otra recta proporcionaría un error cuadrático medio superior al anterior.

Todos estos resultados permiten plantearnos el problema de la **bondad de ajuste**. La recta de regresión de Y sobre X es la mejor explicación lineal de Y a partir de X, pero ¿es buena dicha explicación? Podemos evaluar la bondad de la regresión por el *tamaño* de los errores. Teniendo en cuenta que

$$0 \leq Var(\varepsilon) \leq Var(Y) ,$$

y que

$$E(\varepsilon) = 0 ,$$

valores de $Var(\varepsilon)$ cercanos a cero indicarían escasa variabilidad de la variable ε alrededor del cero, es decir, errores pequeños. Por contra, valores de $Var(\varepsilon)$ cercanos a la varianza de Y, indicarían mucha variablidad de ε, esto es, errores grandes. Ahora bien, dado que

$$Var(\varepsilon) = Var(Y) \cdot (1 - \rho^2) ,$$

será el valor del cuadrado del coeficiente de correlación el que informe sobre el *tamaño* de los errores y, por tanto, sobre la bondad de la regresión. Estudiemos los valores extremos de ρ^2.

Caso 1: $\rho^2 = 1$

Obviamente, $\rho^2 = 1$ ocurrirá si, y sólo si, $Var(\varepsilon) = 0$. Esto será posible siempre y cuando el error, ε, sea constante, y como $E(\varepsilon) = 0$, concretamente igual a cero: $\varepsilon = 0$, esto es, $p[\varepsilon = 0] = 1$. En otras palabras, no hay errores en la regresión, o lo que es lo mismo,

$$\tilde{Y} = Y ,$$

es decir, la explicación que $\tilde{Y}$ proporciona de Y es *perfecta*. Por tanto, entre X e Y existe una relación lineal exacta que viene expresada por la recta de regresión.

Ilustrando esta situación con una nube de puntos referida a una variable aleatoria bidimensional discreta, vemos que todos los puntos de la nube están en línea recta y por todos ellos pasa la recta de regresión (figura 7.5).

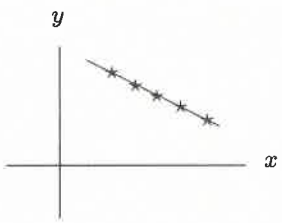


FIG. 7.5.

En resumen, si $\rho^2 = 1$, la recta de regresión de Y sobre X proporciona una explicación de Y perfecta, sin errores, dado que entre las variables X e Y existe

una relación lineal exacta, determinista. La proposición recíproca es igualmente cierta: si existe una relación lineal entre X e Y, el coeficiente de correlación al cuadrado valdrá uno.

Caso 2: $\rho^2 = 0$

La situación $\rho^2 = 0$ ocurrirá si, y sólo si, $Var(\varepsilon) = Var(Y)$, es decir, cuando la variable error alcance su máxima variabilidad. El gran *tamaño* de los errores pone en evidencia que la recta de regresión de Y sobre X no explica nada de la variabilidad de Y.

Entonces, $Var(\tilde{Y}) = 0$ y, como $E(\tilde{Y}) = E(Y)$, la variable Y *teórico*, $\tilde{Y}$, será una constante, en concreto,

$$\tilde{Y} = E(Y) ,$$

o, expresado como relación entre variables matemáticas,

$$y = E(Y) .$$

Es decir, sea cual sea el valor de X, la recta de regresión siempre proporciona el mismo valor. Todo ello se debe a que no existe ningún atisbo de relación lineal entre las variables, de ahí que estén incorrelacionadas. La figura 7.6 ilustra esta situación.

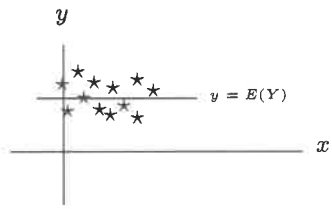


FIG. 7.6.

Obviamente, los dos casos anteriores se presentan raramente en la práctica. Lo normal es que nos encontremos ante situaciones intermedias que interpretaremos por aproximación. Así, valores altos de ρ^2 (por convenio, se suelen considerar valores de ρ^2 mayores o iguales que 0.64) indican relaciones más o menos lineales entre las variables y, por tanto, la recta de regresión de Y sobre X proporciona, en este caso, una explicación aceptable de Y, es decir, con pocos errores.

Por el contrario, valores de ρ^2 próximos a cero ponen en evidencia una escasa relación lineal entre las variables y, en consecuencia, la inconveniente explicación que de Y proporciona la recta de regresión.

La misma discusión sobre la bondad de ajuste que hemos hecho para la recta de regresión de Y sobre X, la podemos trasladar a la de X sobre Y. Así, si ε' es el error cometido en la regresión, es decir,

$$\varepsilon' = X - \tilde{X} ,$$

obtendríamos, por un lado,

$$Var(X) = Var(\tilde{X}) + Var(\varepsilon') \,,$$

y, por otro,

$$Var(\varepsilon') = Var(X) \cdot (1 - \rho^2) \,.$$

De nuevo, será el cuadrado del coeficiente de correlación el que permita discutir sobre la bondad del ajuste de la regresión de X sobre Y. En otras palabras, valores altos de ρ^2 indican *buena* regresión de Y sobre X y de X sobre Y, mientras que valores próximos a cero indican la inconveniencia de ambas regresiones. Evidentemente, si X e Y están incorrelacionadas, la recta de regresión de X sobre Y será

$$\tilde{X} = E(X) \,,$$

o, como recta matemática,

$$x = E(X) \,,$$

situación que se muestra en la figura 7.7.

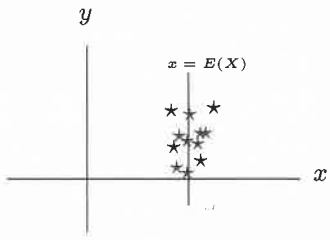


FIG. 7.7.

En todos los comentarios acerca de la bondad del ajuste, hemos prescindido del signo del coeficiente de correlación, al considerarlo siempre elevado al cuadrado; además, hemos analizado las rectas de regresión separadamente. A continuación, incorporaremos ambos hechos al estudio de la regresión. Veamos, previamente, dos propiedades sobre las pendientes de las rectas de regresión.

P1. La covarianza, los coeficientes de regresión, las pendientes de las rectas de regresión y el coeficiente de correlación tienen el mismo signo.

Demostración. Obviamente, las características

$$Cov(X,Y), \quad b_{Y\,|\,X} = \frac{Cov(X,Y)}{Var(X)} \,, \quad b_{X\,|\,Y} = \frac{Cov(X,Y)}{Var(Y)} \,,$$

$$p_{Y\,|\,X} = \frac{Cov(X,Y)}{Var(X)} \quad \text{y} \quad p_{X\,|\,Y} = \frac{Var(Y)}{Cov(X,Y)} \,,$$

tienen todas el mismo signo, que es signo de la covarianza. (Si ésta se anulase, se anularían también el resto de las características, salvo $p_{X\,|\,Y}$, que no estaría definida.) ♣

De aquí, deducimos que las rectas de regresión son ambas crecientes (pendientes positivas) o decrecientes (pendientes negativas).

P2. Las pendientes de las rectas de regresión verifican que

$$|p_{X\,|\,Y}| \geq |p_{Y\,|\,X}| \,.$$

Demostración. Efectivamente,

$$b_{Y\,|\,X} \cdot b_{X\,|\,Y} = \frac{Cov(X,Y)}{Var(X)} \cdot \frac{Cov(X,Y)}{Var(Y)} = \rho^2 \leq 1 \,.$$

Entonces, si los coeficientes de regresión son positivos,

$$b_{Y\,|\,X} \leq \frac{1}{b_{X\,|\,Y}} \,,$$

y, por tanto,

$$p_{Y\,|\,X} \leq p_{X\,|\,Y} \,.$$

Es decir, cuando las rectas de regresión crecen, la de X sobre Y crece más acusadamente que la de Y sobre X.

Por otro lado, si los coeficientes de regresión son negativos,

$$b_{Y\,|\,X} \geq \frac{1}{b_{X\,|\,Y}} \,,$$

esto es,

$$p_{Y\,|\,X} \geq p_{X\,|\,Y} \,.$$

Consecuentemente, cuando las rectas son decrecientes, es más rápido el decrecimiento de la recta de regresión de X sobre Y. ♣

Todo lo anterior conduce a analizar e interpretar el valor del coeficiente de correlación y su signo.

Caso 1: Valores de ρ próximos a 1 ($\rho \geq 0.8$)

Existe una cierta relación lineal y creciente o directa entre las variables (correlación positiva), proporcionando la recta de regresión de Y sobre X y la de X sobre Y una aceptable explicación de Y y de X, respectivamente.

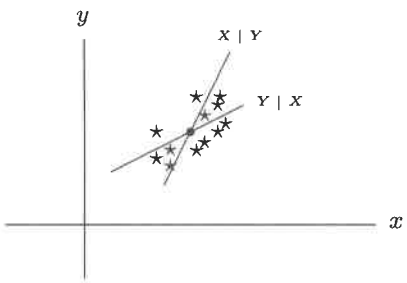


FIG. 7.8.

Un caso particular se plantea cuando $\rho = 1$. Entonces, entre X e Y existe una relación lineal creciente exacta. Las pendientes, positivas, de las rectas de regresión son iguales ya que, como

$$b_{Y \mid X} \cdot b_{X \mid Y} = \rho^2 = 1 \, ,$$

se verifica que

$$p_{Y \mid X} = b_{Y \mid X} = \frac{1}{b_{X \mid Y}} = p_{X \mid Y} \, .$$

Entonces, las rectas coinciden, pues tienen la misma pendiente y un punto común, $(E(X), E(Y))$.

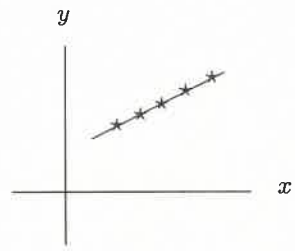


FIG. 7.9.

Caso 2: Valores de ρ próximos a -1 ($\rho \leq -0.8$)

De igual forma, ahora existirá una cierta relación lineal y decreciente o inversa entre las variables (correlación negativa), proporcionando las rectas de regresión explicaciones aceptables.

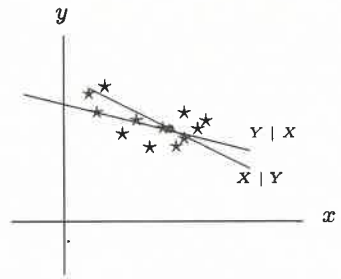


FIG. 7.10.

Cuando $\rho = -1$, entre X e Y hay una relación lineal decreciente exacta, relación lineal que describen las dos rectas de regresión que, al igual que para $\rho = 1$, coinciden, como se muestra en la figura 7.11.

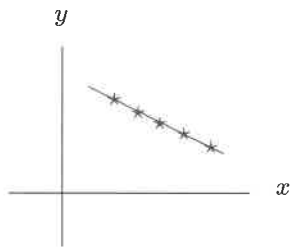


FIG. 7.11.

Caso 3: X e Y están incorrelacionadas ($\rho = 0$)

Como ya se ha dicho, no hay ningún indicio de linealidad entre X e Y, no teniendo ningún sentido la explicación que proporcionan las rectas de regresión. Las dos rectas de regresión son perpendiculares, cortándose en $(E(X), E(Y))$, centro de gravedad de la nube de puntos (figura 7.12).

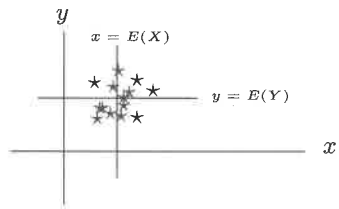


FIG. 7.12.

Por último, veamos dos expresiones del cuadrado del coeficiente de correlación que permiten una interpretación habitual de sus valores. De la descomposición de la varianza deducimos:

$$\begin{aligned} Var(\varepsilon) &= Var(Y) - Var(\tilde{Y}) = \\ &= Var(Y) \cdot \left[1 - \frac{Var(\tilde{Y})}{Var(Y)}\right] . \end{aligned}$$

Pero, por otro lado, sabemos que

$$Var(\varepsilon) = Var(Y) \cdot (1 - \rho^2) .$$

De aquí,

$$\rho^2 = \frac{Var(\tilde{Y})}{Var(Y)} .$$

Haciendo lo mismo para la regresión de X sobre Y, obtendríamos:

$$\rho^2 = \frac{Var(\tilde{X})}{Var(X)} .$$

En este sentido, ρ^2 representa la proporción entre la varianza explicada por la regresión y la varianza a explicar. Así, por ejemplo, un $\rho^2 = 0.85$, indicaría que la regresión explica el 85 % de la variabilidad de la variable que queremos explicar.

La regresión no sólo sirve para explicar una variable a partir de otra, explicación que no tiene por qué ser obligatoriamente causal. También sirve para predecir el valor de una variable cuando la otra ha tomado un determinado valor. Para que esta predicción sea fiable, hay que exigir que la regresión sea aceptable (ρ^2 cercano a la unidad).

Veamos a continuación dos ejemplos, el primero para variables discretas y el segundo para continuas, que ilustran la regresión lineal.

Ejemplo 7.7. Tratemos de explicar la variable Y, *número de bolas negras antes de extraer la primera bola blanca*, mediante una función lineal de X, *número de bolas blancas extraídas*, al hacer tres extracciones sin devolución de una urna con tres bolas blancas y dos bolas negras, variables cuyas distribuciones conjuntas y marginales, calculadas en el ejemplo 7.4, vienen recogidas en la siguiente tabla:

$X\backslash Y$	0	1	2	
1	1/10	1/10	1/10	3/10
2	4/10	2/10	0	6/10
3	1/10	0	0	1/10
	6/10	3/10	1/10	1

Tendremos que hallar la recta de regresión de Y sobre X, esto es,

$$\tilde{Y} - E(Y) = \frac{Cov(X,Y)}{Var(X)} \cdot [X - E(X)] ,$$

para lo cual calcularemos previamente las esperanzas y varianzas de X y de Y, así como la covarianza entre ambas. Por un lado,

$$E(X) = 1.8$$

y

$$E(Y) = 0.5 .$$

Por otro, y como $E(X^2) = 3.6$ y $E(Y^2) = 0.7$, las varianzas resultan:

$$Var(X) = E(X^2) - [E(X)]^2 = 0.36$$

y

$$Var(Y) = E(Y^2) - [E(Y)]^2 = 0.45 .$$

Por último, calculemos la covarianza entre X e Y; como

$$E(X \cdot Y) = 0.7\,,$$

entonces,

$$Cov(X,Y) = E(X \cdot Y) - E(X)E(Y) = -0.2\,.$$

Por tanto, la mejor explicación lineal de Y a partir de X será:

$$\tilde{Y} - 0.5 = -\frac{0.2}{0.36}(X - 1.8)\,,$$

es decir,

$$\tilde{Y} = -\frac{5}{9}X + \frac{3}{2}\,.$$

Luego, la distribución de la variable $\tilde{Y}$ será:

$$\tilde{Y}: \left| \begin{array}{ll} \frac{17}{18} & p\left[\tilde{Y} = \frac{17}{18}\right] = p[X=1] = \frac{3}{10} \\ \frac{7}{18} & p\left[\tilde{Y} = \frac{7}{18}\right] = p[X=2] = \frac{6}{10} \\ -\frac{3}{18} & p\left[\tilde{Y} = \frac{-3}{18}\right] = p[X=3] = \frac{1}{10} \end{array} \right.$$

de donde deducimos que $E(\tilde{Y}) = 0.5$ y $Var(\tilde{Y}) = \frac{1}{9}$.

Por su parte, la variable ε,

$$\varepsilon = Y - \tilde{Y} = Y + \frac{5}{9}X - \frac{3}{2}\,,$$

tendrá la siguiente distribución:

$$\varepsilon: \left| \begin{array}{ll} -\frac{17}{18} & p\left[\varepsilon = -\frac{17}{18}\right] = p[X=1, Y=0] = \frac{1}{10} \\ \frac{1}{18} & p\left[\varepsilon = \frac{1}{18}\right] = p[X=1, Y=1] = \frac{1}{10} \\ \frac{19}{18} & p\left[\varepsilon = \frac{19}{18}\right] = p[X=1, Y=2] = \frac{1}{10} \\ -\frac{7}{18} & p\left[\varepsilon = -\frac{7}{18}\right] = p[X=2, Y=0] = \frac{4}{10} \\ \frac{11}{18} & p\left[\varepsilon = \frac{11}{18}\right] = p[X=2, Y=1] = \frac{2}{10} \\ \frac{3}{18} & p\left[\varepsilon = \frac{3}{18}\right] = p[X=3, Y=0] = \frac{1}{10} \end{array} \right.$$

siendo su esperanza nula, $E(\varepsilon) = 0$, como ya sabemos.

El error cuadrático esperado será:

$$E(\varepsilon^2) = \sum_i \varepsilon_i^2 p[\varepsilon = \varepsilon_i] = \frac{61}{180} .$$

Compruebe el lector que

$$Var(Y) = Var(\tilde{Y}) + Var(\varepsilon) .$$

Dibujemos la gráfica de la recta de regresión

$$y = -\frac{5}{9}x + \frac{3}{2} ,$$

sobre un sistema de ejes cartesianos, en el que se ha representado la nube de puntos (figura 7.13).

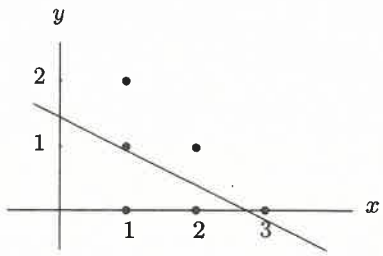


FIG. 7.13.

Las longitudes de cada punto de la nube a la recta (segmentos perpendiculares al eje de abscisas) representan los valores de la variable ε.

Asimismo, podemos calcular la recta de regresión de X sobre Y,

$$\tilde{X} - E(X) = \frac{Cov(X,Y)}{Var(Y)}[Y - E(Y)] ,$$

que, sustituyendo, resulta:

$$\tilde{X} - 1.8 = -\frac{0.2}{0.45}(Y - 0.5) ,$$

es decir,

$$\tilde{X} = -\frac{4}{9}Y + \frac{91}{45} .$$

Representemos conjuntamente sobre un sistema de ejes cartesianos la recta de regresión de Y sobre X,

$$y = -\frac{5}{9}x + \frac{3}{2} ,$$

y la de X sobre Y,

$$x = -\frac{4}{9}y + \frac{91}{45} ,$$

que, expresada en la forma habitual de las rectas, es

$$y = -\frac{9}{4}x + \frac{91}{20} .$$

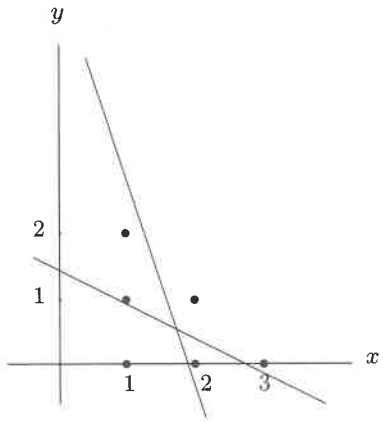


FIG. 7.14.

Estudiemos la bondad del ajuste mediante el coeficiente de correlación:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-0.2}{\sqrt{0.36}\sqrt{0.45}} = -0.497 \, .$$

Por tanto, entre X e Y no existe casi relación lineal y las rectas de regresión proporcionan una mala explicación de cada una de las variables. De hecho, como $\rho^2 = 0.257$, la regresión explica tan sólo el 25.7 % de la variabilidad de la variable.

Aunque lo anterior hace que desestimemos cualquier predicción, ilustremos la técnica que hay que utilizar en el caso en que esto fuera posible. Si sabemos que se extrajeron dos bolas blancas, $X = 2$, el valor de $\tilde{Y}$ proporcionado por la regresión será:

$$\tilde{Y} = -\frac{5}{9} \cdot 2 + \frac{3}{2} = 0.39 \, ,$$

es decir, se predice que saldrán 0.39 bolas negras antes de extraer la primera blanca, cuando se sabe que se extrajeron dos bolas blancas.

Si, por el contrario, sabemos que salió una bola negra antes de la primera blanca, $Y = 1$, el valor de $\tilde{X}$ será:

$$\tilde{X} = -\frac{4}{9} \cdot 1 + \frac{91}{45} = 1.58 \, ,$$

es decir, se prevé que se extrajeron 1.65 bolas blancas. ◁

Ejemplo 7.8. Se pretende explicar la variable Y, *consumo en bienes perecederos*, mediante una función lineal de la variable X, *renta anual disponible* de las familias españolas, medidas ambas variables en millones de pesetas. El comportamiento de estas variables viene modelizado por la siguiente función de densidad, ya estudiada en los ejemplos 7.5 y 7.6:

$$f_{(X,Y)}(x,y) = \begin{cases} e^{-x} & 0 < y < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

Para explicar la variable Y mediante la variable X, hallemos la recta de regresión de Y sobre X:

$$\tilde{Y} - E(Y) = \frac{Cov(X,Y)}{Var(X)} \cdot [X - E(X)] \, .$$

Las densidades marginales de X y de Y son

$$f_X(x) = \int_{y=0}^{x} e^{-x} dy = \begin{cases} xe^{-x} & 0 < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

y

$$f_Y(y) = \int_{x=y}^{\infty} e^{-x} dx = \begin{cases} e^{-y} & 0 < y < \infty \\ 0 & \text{en el resto} \end{cases}$$

Entonces, las esperanzas resultan

$$E(X) = \int_{x=0}^{\infty} x \cdot xe^{-x} dx = 2$$

y

$$E(Y) = \int_{y=0}^{\infty} y \cdot e^{-y} dy = 1 .$$

Para obtener las varianzas, calculemos primero los momentos no centrales de segundo orden:

$$E(X^2) = \int_{x=0}^{\infty} x^2 \cdot xe^{-x} dx = 6$$

y

$$E(Y^2) = \int_{y=0}^{\infty} y^2 \cdot e^{-y} dy = 2 .$$

Por tanto,

$$Var(X) = E(X^2) - [E(X)]^2 = 6 - 2^2 = 2$$

y

$$Var(Y) = E(Y^2) - [E(Y)]^2 = 2 - 1^2 = 1 .$$

Por último,

$$E(X \cdot Y) = \int_{x=0}^{\infty} \int_{y=0}^{x} xy \cdot e^{-x} dy dx = 3 ,$$

de donde la covarianza es igual a

$$Cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 3 - 2 \cdot 1 = 1 .$$

Estamos ya en disposición de dar la recta de regresión de Y sobre X:

$$\tilde{Y} - 1 = \frac{1}{2} \cdot (X - 2) ,$$

es decir,

$$\tilde{Y} = \frac{X}{2} .$$

Esta nueva variable, $\tilde{Y}$, cuya distribución, que no necesitaremos, se calcularía con facilidad, presenta las siguientes características:

$$E(\tilde{Y}) = E\left(\frac{X}{2}\right) = \frac{E(X)}{2} = 1$$

y

$$Var(\tilde{Y}) = Var\left(\frac{X}{2}\right) = \frac{Var(X)}{4} = \frac{1}{2} .$$

Por su parte, para la variable ε,

$$\varepsilon = Y - \tilde{Y} = Y - \frac{X}{2} ,$$

de la cual también se podrá calcular su distribución, tenemos:

$$E(\varepsilon) = E(Y) - E(\tilde{Y}) = 0$$

y

$$Var(\varepsilon) = Var\left(Y - \frac{X}{2}\right) = Var(Y) + \frac{1}{4}Var(X) - Cov(X,Y) = \frac{1}{2} .$$

Compruebe el lector que, de nuevo, se verifica la descomposición de la varianza.

Volviendo a la recta de regresión obtenida,

$$y = \frac{x}{2} ,$$

ésta explica el *consumo en bienes perecederos* a partir de la *renta disponible* (sobre la validez de la explicación hablaremos posteriormente), en el sentido que las familias estudiadas dedican aproximadamente la mitad de su renta a la compra de bienes perecederos.

Asimismo, la recta de regresión permite predecir el consumo de una familia con un cierto nivel de renta. De esta forma, una familia cuya renta sea de 2.5 millones de pesetas, tendrá un consumo teórico, $\tilde{Y}$, de 1.25 millones de pesetas.

Aunque desde el punto de vista de la teoría económica carezca de sentido explicar la renta a partir del consumo, es posible calcular la recta de regresión de X sobre Y. Así, sustituyendo los valores obtenidos en la expresión de la recta de regresión de X sobre Y resultará:

$$\tilde{X} - 2 = \frac{1}{1} \cdot (Y - 1) ,$$

es decir,

$$\tilde{X} = Y + 1 .$$

Representemos, sobre un sistema de ejes cartesianos, las dos rectas de regresión,

$$y = \frac{x}{2} \qquad \text{y} \qquad x = y + 1 .$$

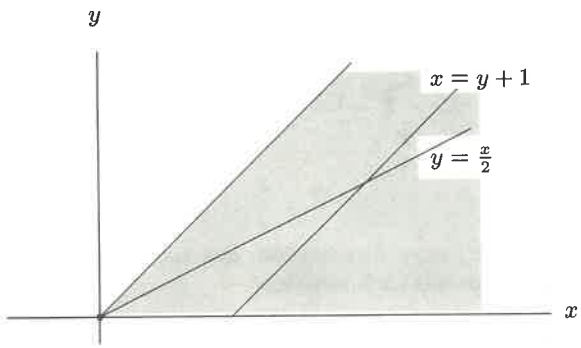


FIG. 7.15.

Estudiemos la bondad del ajuste, calculando el coeficiente de correlación:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{1}{\sqrt{2} \cdot 1} = 0.71 \, .$$

De aquí deducimos que existe una leve relación lineal creciente entre las variables; la recta de regresión de Y sobre X proporciona una insuficiente explicación de Y (lo mismo podíamos decir de la recta de regresión de X sobre Y, con la salvedad hecha anteriormente sobre la inconsistencia teórica de la misma). Teniendo en cuenta que $\rho^2 = 0.5$, la regresión explica la mitad de la variabilidad de la variable. ◁

7.4.2 Regresión general

En el subapartado anterior hemos visto un caso particular (el más importante) de la regresión con restricciones: la regresión lineal. Veamos ahora qué ocurre cuando no ponemos ninguna restricción a la función que va a explicar una variable a partir de la otra.

Sea (X, Y) una variable aleatoria bidimensional. Tratemos de encontrar una función cualquiera, g, que proporcione la mejor explicación de Y a partir de X, de acuerdo con el criterio de los mínimos cuadrados.

Siendo $\tilde{Y} = g(X)$, la variable Y *teórica*, y $\varepsilon = Y - \tilde{Y}$ la variable error, buscaremos g bajo la siguiente condición:

$$\min_g E(\varepsilon^2) = \min_g E[(Y - g(X))^2] \, .$$

Como g no es una función que quede determinada por una serie de parámetros, como ocurría en la regresión lineal, debemos seguir un razonamiento singular para encontrarla.

Según vimos en el apartado 7.2,

$$E[E(h(X,Y) \mid X = x)] = E[h(X,Y)] \, ,$$

con lo cual, deducimos que

$$E(\varepsilon^2) = E[(Y - g(X))^2] = E[E((Y - g(X))^2 \mid X = x)] \, .$$

Teniendo en cuenta el teorema de KÖNIG[4],

$$E[(Y - g(X))^2 \mid X = x]$$

se hace mínima cuando $g(x)$ es la esperanza, es decir,

$$g(x) = E(Y \mid X = x) \, .$$

4. El teorema de KÖNIG (propiedad 3 del subapartado 6.2.2) afirma que $E[(Y - C^2)]$ alcanza el mínimo valor cuando $C = E(Y)$, siendo la varianza, por tanto, ese mínimo valor. Considerando que la esperanza condicionada, esperanza de una distribución (condicionada), verifica todas las propiedades y resultados generales de la esperanza de una variable, también verificará dicho teorema.

Por tanto,

$$E[(Y - g(X))^2 \mid X = x] \geq E[(Y - E(Y \mid X = x))^2 \mid X = x] .$$

De aquí,

$$E(\varepsilon^2) = E[E((Y - g(X))^2 \mid X = x)] \geq E[E((Y - E(Y \mid X = x))^2 \mid X = x)] .$$

En consecuencia,

$$g(x) = E(Y \mid X = x)$$

es la función que hace mínima la esperanza de los errores al cuadrado. Entonces[5]

$$\tilde{Y} = g(X) = E(Y \mid X = x)$$

proporciona la mejor explicación de Y a partir de X, recibiendo el nombre de **función o curva general de regresión de Y sobre X**. Cuando nos refiramos, sin más, a la relación matemática, utilizaremos la notación habitual entre funciones, esto es,

$$y = E(Y \mid X = x) .$$

Podemos realizar el mismo razonamiento cuando se trata de explicar X a partir de Y. En este caso, siendo $\tilde{X} = g(Y)$ y $\varepsilon' = X - \tilde{X}$, buscaremos la función, g, que minimice el error cuadrático medio, es decir,

$$\min_g E\left(\varepsilon'^2\right) = \min_g E[(X - g(Y))^2] ,$$

obteniendo, en este caso,

$$\tilde{X} = g(Y) = E(X \mid Y = y)$$

función o curva general de regresión de X sobre Y, que escrita como relación entre variables matemáticas será

$$x = E(X \mid Y = y) .$$

Obsérvese que los resultados obtenidos son razonables. Así, cuando X toma el valor x, la regresión estima como valor de Y su valor esperado, $E(Y \mid X = x)$.

Como ya adelantamos, la regresión general presenta el inconveniente de proporcionar funciones de uso poco habitual, con la consiguiente dificultad de manejo e interpretación. Ahora bien, en ciertas situaciones, la regresión general proporciona funciones sencillas, por ejemplo, lineales. Cuando esto ocurre, es decir, cuando la mejor explicación sin restricciones es una función lineal, una recta, obviamente coincidirá con la recta de regresión.

Al realizar la regresión general de Y sobre X, partimos de dos variables X e Y, de las que conocemos, al menos, μ_X, μ_Y, σ_X^2, σ_Y^2 y σ_{XY}. En el proceso de

5. Al sustituir x por X, debiéramos cambiar $E(Y \mid X = x)$ por $E(Y \mid X)$, de acuerdo con la notación alternativa que describimos en el apartado 7.2. No lo haremos para no complicar las expresiones al lector, aunque debe entenderse que $\tilde{Y}$ es una función de X y no de x.

regresión se generan dos nuevas variables: $\tilde{Y}$ y ε, es decir, Y *teórico* y el error de la regresión. Estudiemos sus principales características.

La esperanza de $\tilde{Y} = E(Y \mid X = x)$ es

$$E(\tilde{Y}) = E[E(Y \mid X = x)] = E(Y) ,$$

y su varianza,

$$Var(\tilde{Y}) = Var[E(Y \mid X = x)] = Var(Y) - E[Var(Y \mid X = x)] ,$$

de acuerdo con la segunda propiedad del apartado 7.3.

Por otro lado,

$$E(\varepsilon) = E(Y - \tilde{Y}) = E(Y) - E(\tilde{Y}) = 0$$

y

$$\begin{aligned} Var(\varepsilon) &= E(\varepsilon^2) = E[(Y - \tilde{Y})^2] = E[E(Y - E(Y \mid X = x))^2] = \\ &= E[Var(Y \mid X = x)] = Var(Y) - Var[E(Y \mid X = x)] = \\ &= Var(Y) - Var(\tilde{Y}) . \end{aligned}$$

Por tanto,

$$Var(Y) = Var(\tilde{Y}) + Var(\varepsilon) ,$$

expresión que representa la **descomposición de la varianza**, es decir, proporciona la varianza a explicar, $Var(Y)$, como suma de la varianza explicada por la regresión, $Var(\tilde{Y})$, y la varianza no explicada o residual, $Var(\varepsilon)$. En la medida en que uno de los dos sumandos aporte más o menos a la varianza total nos encontraremos ante una *buena* regresión ($Var(\tilde{Y})$ grande, $Var(\varepsilon)$ pequeña) o ante una *mala* regresión ($Var(\tilde{Y})$ pequeña, $Var(\varepsilon)$ grande). Por otra parte,

$$Var(\varepsilon) = Var(Y) - Var(\tilde{Y}) = Var(Y) \cdot \left[1 - \frac{Var(\tilde{Y})}{Var(Y)}\right] .$$

Por analogía con la regresión lineal, llamaremos **razón de correlación de Y sobre X** a la característica

$$\eta^2_{Y \mid X} = \frac{Var(\tilde{Y})}{Var(Y)} ,$$

que tiene en la regresión general el mismo papel que representaba el coeficiente de correlación al cuadrado, ρ^2, en la regresión lineal. Al igual que éste, la razón de correlación está comprendida entre cero y uno, es decir,

$$0 \leq \eta^2_{Y \mid X} \leq 1 ,$$

dado que es un cociente de números positivos, cuyo numerador es siempre menor o igual que el denominador.

Asimismo, si X e Y son independientes, la razón de correlación de Y sobre X se anula, pues, como se vio en la segunda propiedad del apartado 7.2, la independencia de X e Y implica que

$$\tilde{Y} = E(Y \mid X = x) = E(Y)$$

y, por tanto,

$$Var(\tilde{Y}) = 0 .$$

Todo lo anterior permite plantearse el problema de la **bondad de ajuste**. La curva de regresión de Y sobre X proporciona la mejor explicación de Y a partir de X, pero ¿en qué medida es satisfactoria esta explicación? Al igual que para la regresión lineal, la contestación a esta pregunta viene dada por el *tamaño* de los errores. Como sabemos,

$$0 \leq Var(\varepsilon) \leq Var(Y)$$

y

$$E(\varepsilon) = 0 .$$

Consecuentemente, valores de la varianza residual cercanos a cero indicarán *errores pequeños*, mientras que valores cercanos a $Var(Y)$ indicarán *errores grandes*. El tamaño de la varianza residual está determinado por la razón de correlación, puesto que

$$Var(\varepsilon) = Var(Y) \cdot (1 - \eta^2_{Y \mid X}) .$$

Caso 1: $\eta^2_{Y \mid X} = 1$

Si $\eta^2_{Y \mid X} = 1$, la varianza residual valdrá cero y, por tanto, el error ε será una constante. Como ε es una variable de media cero,

$$\varepsilon = 0 .$$

Ahora bien, $\varepsilon = Y - \tilde{Y}$, por lo cual

$$Y = \tilde{Y} = E(Y \mid X = x) .$$

Decimos, entonces, que la regresión es perfecta, en el sentido de que la función o curva general de regresión proporciona exactamente la variable Y que tratamos de explicar. Esto ocurre porque, cuando $\eta^2_{Y \mid X} = 1$, existe una relación funcional entre X e Y, relación que la curva general de regresión describe. En efecto, si $\eta^2_{Y \mid X} = 1$, la varianza residual se anula, esto es,

$$Var(\varepsilon) = E[Var(Y|X = x)] = 0 .$$

Pero $Var(Y|X = x)$ es una variable aleatoria no negativa función de X, por lo que, si su esperanza es igual a cero, valdrá cero con probabilidad igual a la unidad, es decir,

$$p[x \in \mathbb{R} \mid Var(Y|X = x) = 0] = 1 .$$

Ahora bien, las variables con varianza nula son variables degeneradas (constantes), lo que permite concluir que a cada valor, x, de la variable aleatoria X,

le corresponde un único valor de la variable Y (salvo, tal vez, a un conjunto de valores de probabilidad nula). En definitiva, Y es una función de la variable X, o al menos esto ocurre con probabilidad igual a la unidad.

La figura 7.16 muestra esta situación para una variable aleatoria bidimensional discreta: la curva de regresión de Y sobre X pasa por todos los puntos de la nube.

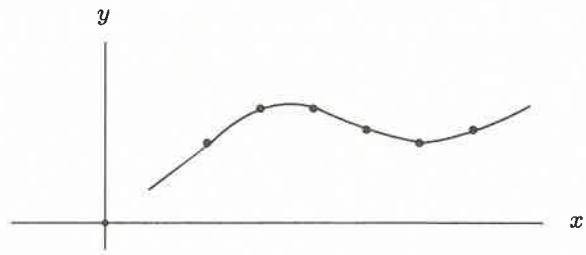


FIG. 7.16.

Cuando $\eta^2_{Y\,|\,X}$ toma valores cercanos a la unidad (por convenio, $\eta^2_{Y\,|\,X} \geq 0.64$) interpretaremos todo lo anterior por aproximación y hablaremos de regresiones más o menos aceptables.

Caso 2: $\eta^2_{Y\,|\,X} = 0$

Si $\eta^2_{Y\,|\,X} = 0$, la varianza residual tomará el mayor valor de los posibles, $Var(\varepsilon) = Var(Y)$, de donde se deduce que $Var(\tilde{Y}) = 0$. Como $E(\tilde{Y}) = E(Y)$, entonces,

$$\tilde{Y} = E(Y)\,,$$

esto es, nos encontramos ante la misma situación que cuando ρ era igual a cero (véase la figura 7.6).

En este caso, la explicación que proporciona la curva de regresión es totalmente inadecuada al no existir atisbos de una relación funcional que proporcione Y a partir de X.

Toda la discusión anterior sobre la bondad del ajuste de la regresión general de Y sobre X se puede trasladar a la regresión de X sobre Y. Ahora, si llamamos ε' al error de la regresión,

$$Var(X) = Var(\tilde{X}) + Var(\varepsilon')\,,$$

o, lo que es lo mismo,

$$Var(\varepsilon') = Var(X) \cdot (1 - \eta^2_{X\,|\,Y})\,,$$

donde $\eta^2_{X\,|\,Y}$, **razón de correlación de X sobre Y**, se define como

$$\eta^2_{X\,|\,Y} = \frac{Var(\tilde{X})}{Var(X)}\,.$$

Las mismas consideraciones que hicimos a partir de los valores de $\eta^2_{Y\mid X}$ para la regresión de Y sobre X, son trasladables a $\eta^2_{X\mid Y}$ para la regresión de X sobre Y. Ahora bien, en general,

$$\eta^2_{Y\mid X} \neq \eta^2_{X\mid Y} ,$$

por lo cual la bondad del ajuste de Y sobre X no tiene por qué coincidir con la de X sobre Y. Esto es, podemos encontrar, por ejemplo, una mala explicación de Y a partir de X ($\eta^2_{Y\mid X}$ próxima a cero) y, en cambio, una buena explicación de X a partir de Y ($\eta^2_{X\mid Y}$ próxima a uno).

Las razones de correlación,

$$\eta^2_{Y\mid X} = \frac{Var(\tilde{Y})}{Var(Y)}$$

y

$$\eta^2_{X\mid Y} = \frac{Var(\tilde{X})}{Var(X)} ,$$

pueden expresarse como el porcentaje de variabilidad explicada sobre la variabilidad a explicar. En este sentido, como la regresión general (sin restricciones) proporciona una mejor explicación que la regresión lineal (con restricciones) parece lógico que las razones de correlación tomen valores mayores o iguales que el cuadrado del coeficiente de correlación:

$$\eta^2_{Y\mid X} \geq \rho^2 \qquad \text{y} \qquad \eta^2_{X\mid Y} \geq \rho^2 .$$

La demostración es inmediata. Recordemos la descomposición de la varianza en la regresión lineal, que ahora denotaremos con el subíndice L,

$$Var(Y) = Var(\tilde{Y}_L) + Var(\varepsilon_L) ,$$

y la descomposición análoga en la regresión general que denotaremos con el subíndice G,

$$Var(Y) = Var(\tilde{Y}_G) + Var(\varepsilon_G) .$$

Como en la regresión lineal ponemos restricciones a la hora de buscar la recta de regresión, cometeremos mayor error cuadrático medio que cuando no exigimos ningún tipo de restricción. Entonces,

$$E(\varepsilon^2_L) \geq E(\varepsilon^2_G) .$$

Pero, dado que $E(\varepsilon_L) = E(\varepsilon_G) = 0$, las varianzas residuales de la regresión lineal y general verifican:

$$Var(\varepsilon_L) \geq Var(\varepsilon_G) ,$$

y, en consecuencia,

$$Var(\tilde{Y}_L) \leq Var(\tilde{Y}_G) ,$$

de donde deducimos que

$$\rho^2 = \frac{Var(\tilde{Y}_L)}{Var(Y)} \leq \frac{Var(\tilde{Y}_G)}{Var(Y)} = \eta^2_{Y\mid X} .$$

De la misma forma demostraríamos que $\eta^2_{X \mid Y} \geq \rho^2$ para la regresión de X sobre Y.

En el caso de que la curva de regresión de Y sobre X sea una recta,

$$\tilde{Y}_L = \tilde{Y}_G \,,$$

siguiendo el razonamiento anterior deduciríamos que

$$\rho^2 = \eta^2_{Y \mid X} \,.$$

Lo mismo sucederá cuando la curva de regresión de X sobre Y sea una recta.

Adviértase que las razones de correlación no admiten una interpretación de su raíz cuadrada como pasaba con ρ (de hecho, el 2 que aparece como superíndice en la razón de correlación sólo busca un paralelismo notacional con ρ^2). Téngase en cuenta que la interpretación de ρ, es decir, de su signo, tiene que ver con las pendientes de las rectas de regresión, o sea, con el sentido, positivo o negativo, de la relación lineal entre las variables. Esta discusión, obviamente, únicamente tiene sentido cuando hablamos de funciones lineales, no así para otro tipo de funciones.

Si bien la complejidad de las curvas de regresión desaconseja su utilización, éstas pueden servir para proporcionar un resumen previo de la nube de puntos. El aspecto funcional de este resumen orientará sobre el tipo de restricción más adecuada que conduzca a un buen ajuste.

Dos ejemplos ilustrarán los conceptos anteriores.

Ejemplo 7.9. En la misma situación que en el ejemplo 7.7, calculemos, ahora, la función o curva general de regresión de Y sobre X, siendo Y la variable *número de bolas negras antes de extraer la primera bola blanca*, y X, *número de bolas blancas extraídas* al hacer tres extracciones sin devolución de una urna con tres bolas blancas y dos bolas negras. La curva de regresión de Y sobre X es

$$\tilde{Y} = E(Y \mid X = x) \,,$$

por lo que debemos calcular la función de probabilidad condicionada.

La probabilidad de Y condicionada a cualquier valor de X es, según los casos,

$$p[Y = y \mid X = 1] = \begin{cases} \frac{1}{3} & \text{si } y = 0 \\ \frac{1}{3} & \text{si } y = 1 \\ \frac{1}{3} & \text{si } y = 2 \end{cases}$$

$$p[Y = y \mid X = 2] = \begin{cases} \frac{2}{3} & \text{si } y = 0 \\ \frac{1}{3} & \text{si } y = 1 \\ 0 & \text{si } y = 2 \end{cases}$$

$$p[Y = y \mid X = 3] = \begin{cases} 1 & \text{si } y = 0 \\ 0 & \text{si } y = 1 \text{ o } y = 2 \end{cases}$$

De aquí obtenemos la esperanza condicionada para cada uno de los valores posibles de X,

$$\begin{aligned} E(Y \mid X = 1) &= 1 \,, \\ E(Y \mid X = 2) &= \frac{1}{3} \quad \text{y} \\ E(Y \mid X = 3) &= 0 \,. \end{aligned}$$

En consecuencia,

$$\tilde{Y} = E(Y \mid X = x) = \begin{cases} 1 & \text{si } x = 1 \\ \frac{1}{3} & \text{si } x = 2 \\ 0 & \text{si } x = 3 \end{cases}$$

Esta función o curva general de regresión es una variable aleatoria función de X, con $E(\tilde{Y}) = 0.5$ y $Var(\tilde{Y}) = \frac{7}{60}$. Por su parte, la distribución de la variable $\varepsilon = Y - \tilde{Y}$ será

ε :	
-1	$p[\varepsilon = -1] = p[X = 1, Y = 0] = \frac{1}{10}$
$-\frac{1}{3}$	$p\left[\varepsilon = -\frac{1}{3}\right] = p[X = 2, Y = 0] = \frac{4}{10}$
0	$p[\varepsilon = 0] = p[X = 1, Y = 1] + p[X = 3, Y = 0] = \frac{2}{10}$
$\frac{2}{3}$	$p\left[\varepsilon = \frac{2}{3}\right] = p[X = 2, Y = 1] = \frac{2}{10}$
1	$p[\varepsilon = 1] = p[X = 1, Y = 2] = \frac{1}{10}$

Por lo tanto,

$$E(\varepsilon) = 0$$

y

$$Var(\varepsilon) = \frac{1}{3} .$$

Puede comprobar el lector que

$$Var(Y) = Var(\tilde{Y}) + Var(\varepsilon) ,$$

aportando la varianza residual mucho más a la varianza total que la varianza explicada por la regresión; es decir, la regresión no va a proporcionar una explicación adecuada de la variable Y.

Si comparamos el error cuadrático medio, o lo que es lo mismo, la varianza residual, de la regresión lineal (ejemplo 7.7) y de la regresión general,

$$E(\varepsilon_L^2) = \frac{61}{180}$$

y

$$E(\varepsilon_G^2) = \frac{1}{3} ,$$

podemos comprobar que, en este caso, la regresión general no aporta mucha más explicación que la lineal.

Por otro lado, la razón de correlación de Y sobre X,

$$\eta^2_{Y \mid X} = \frac{Var(\tilde{Y})}{Var(Y)} = 0.259 ,$$

confirma la escasa validez de la regresión general de Y sobre X: sólo explica el 25.9 % de la variabilidad de Y.

Advierta el lector que la representación gráfica de la función o curva general de regresión de Y sobre X (figura 7.17),

$$y = \begin{cases} 1 & x = 1 \\ \frac{1}{3} & x = 2 \\ 0 & x = 3 \end{cases}$$

no proporciona una curva en el sentido habitual del término, sino que se trata de una serie de puntos, tres en este caso, en el plano.

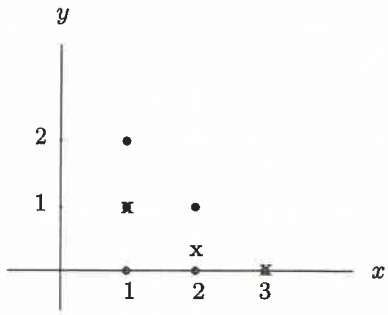


FIG. 7.17.

Las longitudes verticales de cada punto de la nube de puntos a los puntos de la curva de regresión, son los valores de ε.

De modo análogo, puede el lector obtener la función o curva de regresión de X sobre Y, resultando

$$\tilde{X} = E(X \mid Y = y) = \begin{cases} 2 & \text{si } y = 0 \\ \frac{5}{3} & \text{si } y = 1 \\ 1 & \text{si } y = 2 \end{cases}$$

donde $E(\tilde{X}) = 1.8$ y $Var(\tilde{X}) = \frac{7}{75}$. Calculando la correspondiente razón de correlación

$$\eta^2_{X \mid Y} = \frac{Var(\tilde{X})}{Var(X)} = 0.259 ,$$

concluimos que la explicación que de X da la curva de regresión no es aceptable (obsérvese que, en este ejemplo, coincide el valor de las dos razones, algo que no tiene por qué ocurrir.). ◁

Ejemplo 7.10. Queremos explicar, al igual que en el ejemplo 7.8, la variable Y, *consumo en bienes perecederos*, mediante la variable X, *renta anual disponible* de las familias españolas. El comportamiento de estas variables se ajusta a la siguiente función de densidad conjunta:

$$f_{(X,Y)}(x,y) = \begin{cases} e^{-x} & 0 < y < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

Hallemos la función o curva de regresión de Y sobre X,

$$\tilde{Y} = E(Y \mid X = x) .$$

Conocida la función de densidad marginal de X,

$$f_X(x) = \begin{cases} xe^{-x} & 0 < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

se obtiene de inmediato la función de densidad de Y condicionada por $X = x$, para $0 < x < \infty$,

$$f_{(Y \mid X=x)}(y) = \frac{e^{-x}}{xe^{-x}} = \frac{1}{x}\,, \qquad \text{si } 0 < y < x\,,$$

anulándose en el resto.

En consecuencia, la curva de regresión de Y sobre X resulta,

$$E(Y \mid X = x) = \int_{y=0}^{x} y \cdot \frac{1}{x} dy = \frac{x}{2}\,,$$

que, como es una recta, coincide con la recta de regresión.

En este caso, la bondad del ajuste es la misma para la regresión lineal y para la regresión general, pues

$$\eta^2_{Y \mid X} = \rho^2\,.$$

Igualmente, podríamos calcular la curva de regresión de X sobre Y, si bien, como se insistió en el caso lineal, la explicación de X a partir de Y carece de sentido económico. Así, la función de densidad de Y es

$$f_Y(y) = \begin{cases} e^{-y} & 0 < y < \infty \\ 0 & \text{en el resto} \end{cases}$$

De aquí, obtendríamos la función de densidad de X condicionada a $Y = y$, para $0 < y < \infty$,

$$f_{(X \mid Y=y)}(x) = \frac{e^{-x}}{e^{-y}} = e^{-(x-y)}\,, \qquad \text{si } y < x < \infty\,,$$

siendo cero en el resto de los valores de x.

Entonces, la curva de regresión de X sobre Y,

$$E(X \mid Y = y) = \int_{x=y}^{\infty} x \cdot e^{-(x-y)} dx = y + 1\,,$$

también es una recta. Consecuentemente, las curvas generales de regresión coinciden con las rectas de regresión, por lo que

$$\eta^2_{X \mid Y} = \eta^2_{Y \mid X} = \rho^2\,. \triangleleft$$

7.5 Esperanza y matriz de varianzas-covarianzas de una variable aleatoria n-dimensional

Hasta ahora, cuando nos hemos referido a los momentos de una variable aleatoria bidimensional (en general, n-dimensional), los hemos considerado separadamente. En este sentido, dada una variable (X, Y), hemos estudiado, por ejemplo, $E(X)$, $Var(Y)$ o $E(X \cdot Y)$. Introduciremos, ahora, dos nuevos conceptos, los de matriz de esperanzas y matriz de varianzas y covarianzas de una variable aleatoria n-dimensional, conceptos que permitirán englobar en expresiones únicas a los principales momentos marginales y conjuntos de dicha variable aleatoria. Como veremos, esta agrupación simplifica notablemente el cálculo de características de variables aleatorias n-dimensionales y facilita la interpretación de aquéllas.

Definamos, inicialmente, la esperanza de una matriz aleatoria, matriz cuyos elementos son variables aleatorias, aunque no se excluye la posibilidad de que algunas de sus componentes sean degeneradas, es decir, constantes.

Definición 7.9 *La esperanza de una matriz aleatoria es la matriz de las esperanzas de sus elementos. Es decir, si* $\mathbf{Z}$ *es una matriz* $m \times n$ *cuyos elementos son variables aleatorias,*

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \vdots & \vdots \\ Z_{m1} & \cdots & Z_{mn} \end{pmatrix},$$

entonces, la esperanza de $\mathbf{Z}$ *es la siguiente matriz* $m \times n$*:*

$$E(\mathbf{Z}) = \begin{pmatrix} E(Z_{11}) & \cdots & E(Z_{1n}) \\ \vdots & \vdots & \vdots \\ E(Z_{m1}) & \cdots & E(Z_{mn}) \end{pmatrix}.$$

Para simplificar la notación, designaremos las matrices por su término general, esto es, describiremos la matriz $\mathbf{Z}$ anterior como (Z_{ij}).

Como puede verse, la definición 7.9 no introduce ningún concepto, sino que, únicamente, asigna un nombre a la agrupación de las esperanzas de varias variables aleatorias. Su interés está en que es posible utilizar todas las esperanzas en grupo cuando se incluyen en una matriz, como veremos a continuación al estudiar sus propiedades.

P1. Sea $\mathbf{A} = (a_{ij})$ una matriz cuyos elementos, a_{ij}, son constantes; entonces,

$$E(\mathbf{A}) = \mathbf{A}\,.$$

Demostración. Como $E(\mathbf{A})$ es la matriz $(E(a_{ij}))$ y, por otro lado, la esperanza de una constante es ella misma,

$$E(\mathbf{A}) = (E(a_{ij})) = (a_{ij}) = \mathbf{A}\,. \clubsuit$$

P2. Sea $\mathbf{A} = (a_{ij})$ una matriz $p \times m$, cuyos elementos son constantes, y sea $\mathbf{Z} = (Z_{ij})$ una matriz aleatoria $m \times n$; entonces,

$$E(\mathbf{A} \cdot \mathbf{Z}) = \mathbf{A} \cdot E(\mathbf{Z})\,.$$

Demostración. Obviamente, $\mathbf{A} \cdot \mathbf{Z}$ es una matriz $p \times n$ cuyo término general es

$$a_{i1} \cdot Z_{1j} + \cdots + a_{im} \cdot Z_{mj}\,.$$

Entonces, su esperanza, $E(\mathbf{A} \cdot \mathbf{Z})$, tiene por término general

$$E(a_{i1} \cdot Z_{1j} + \cdots + a_{im} \cdot Z_{mj}) = a_{i1} \cdot E(Z_{1j}) + \cdots + a_{im} \cdot E(Z_{mj})\,,$$

que es el término general de la matriz $\mathbf{A} \cdot E(\mathbf{Z})$. En consecuencia,

$$E(\mathbf{A} \cdot \mathbf{Z}) = \mathbf{A} \cdot E(\mathbf{Z}) \; . \; \clubsuit$$

P3. Sea $\mathbf{A} = (a_{ij})$ una matriz $n \times p$, cuyos elementos son constantes, y sea $\mathbf{Z} = (Z_{ij})$ una matriz aleatoria $m \times n$; entonces,

$$E(\mathbf{Z} \cdot \mathbf{A}) = E(\mathbf{Z}) \cdot \mathbf{A} \; .$$

La demostración es idéntica a la anterior.

P4. Sean $\mathbf{Z} = (Z_{ij})$ y $\mathbf{U} = (U_{ij})$ dos matrices aleatorias de dimensión $m \times n$; entonces,

$$E(\mathbf{Z} + \mathbf{U}) = E(\mathbf{Z}) + E(\mathbf{U}) \; .$$

Demostración. El término general de la matriz suma, $\mathbf{Z} + \mathbf{U}$, es $Z_{ij} + U_{ij}$, luego su esperanza será:

$$E(\mathbf{Z} + \mathbf{U}) = (E(Z_{ij} + U_{ij})) = (E(Z_{ij}) + E(U_{ij})) = E(\mathbf{Z}) + E(\mathbf{U}) \; . \; \clubsuit$$

Evidentemente, si la matriz $\mathbf{U}$ es una matriz de constantes, el resultado se concreta en la igualdad

$$E(\mathbf{Z} + \mathbf{U}) = E(\mathbf{Z}) + \mathbf{U} \; .$$

Asimismo, esta propiedad también se verifica para la suma de cualquier conjunto de matrices de la misma dimensión.

Las propiedades anteriores son generalizaciones, para matrices aleatorias, de las propiedades que demostramos para la esperanza de variables aleatorias unidimensionales. La primera indica que la esperanza de una matriz constante es ella misma, y las restantes se resumen en la afirmación de que la esperanza de *combinaciones lineales* de matrices aleatorias son las combinaciones lineales de las esperanzas de dichas matrices. Hemos destacado la expresión «combinaciones lineales» porque las matrices de constantes no juegan el mismo papel que las constantes reales (observará el lector que, por ejemplo, las propiedades **2** y **3** no son la misma, sino que demuestran resultados diferentes, ya que $\mathbf{A} \cdot \mathbf{Z}$ no coincide con $\mathbf{Z} \cdot \mathbf{A}$; de hecho, a veces existe el primer producto y ni siquiera existe el segundo).

Si $(X_1, \ldots, X_n)$ es una variable aleatoria n-dimensional, pueden aplicársele la definición y propiedades anteriores, sin más que considerar dicha variable como una matriz aleatoria. Para evitar confusiones, siempre que entendamos una variable aleatoria n-dimensional como una matriz, supondremos que se trata de una matriz columna, esto es, una matriz $n \times 1$.

Definamos, con esta nueva notación, la esperanza de una variable aleatoria n-dimensional.

Definición 7.10 *Sea* $\mathbf{X}$ *una variable aleatoria n-dimensional, esto es,*

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} .$$

Llamamos esperanza de $\mathbf{X}$, $E(\mathbf{X})$, *a la matriz*

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} .$$

Es decir, $E(\mathbf{X})$ es la matriz columna formada por cada una de las esperanzas de las componentes de $\mathbf{X}$.

La siguiente proposición proporciona un importante resultado.

Proposición 7.1 *Sean* $\mathbf{A}$ *y* $\mathbf{B}$ *dos matrices cuyos elementos son números reales de dimensiones* $m \times n$ *y* $m \times 1$, *respectivamente. Sea* $\mathbf{X}$ *una variable aleatoria n-dimensional. Entonces,*

$$E(\mathbf{A} \cdot \mathbf{X} + \mathbf{B}) = \mathbf{A} \cdot E(\mathbf{X}) + \mathbf{B} .$$

Este resultado se deduce como caso particular de las propiedades de la esperanza de una matriz.

Adviértase que este resultado proporciona la esperanza de una variable aleatoria m-dimensional, $\mathbf{Y} = \mathbf{A} \cdot \mathbf{X} + \mathbf{B}$, cuyas componentes, $Y_1, \ldots, Y_m$, son combinación lineal de las componentes de $\mathbf{X}$, $X_1, \ldots, X_n$. Esto es,

$$\begin{aligned} Y_1 &= a_{11}X_1 + \cdots + a_{1n}X_n + b_1 \\ &\vdots \qquad\qquad \vdots \\ Y_m &= a_{m1}X_1 + \cdots + a_{mn}X_n + b_m \end{aligned}$$

La siguiente definición permite agrupar la información sobre las varianzas y covarianzas de las componentes de una variable aleatoria n-dimensional.

Definición 7.11 *Llamamos matriz de varianzas-covarianzas de la variable aleatoria n-dimensional* $\mathbf{X}$ *a*

$$V(\mathbf{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} ,$$

siendo

$$\sigma_{ij} = \begin{cases} Var(X_i) & i = j \\ Cov(X_i, X_j) & i \neq j \end{cases}$$

Ahora bien, el producto de la matriz $(\mathbf{X} - E(\mathbf{X}))$ por su matriz transpuesta, $(\mathbf{X} - E(\mathbf{X}))'$, es la matriz de término general

$$(X_i - E(X_i)) \cdot (X_j - E(X_j)) .$$

Por tanto,

$$E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'] = (E[(X_i - E(X_i)) \cdot (X_j - E(X_j)]) = (\sigma_{ij}) .$$

En definitiva, la matriz $V(\mathbf{X})$ de varianzas-covarianzas puede expresarse como la esperanza de la matriz $(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'$, es decir,

$$V(\mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'] .$$

Veamos unas propiedades de la matriz de varianzas-covarianzas.

P1. La matriz de varianzas-covarianzas de una variable aleatoria $\mathbf{X}$ se puede expresar como

$$V(\mathbf{X}) = E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})(E(\mathbf{X}))' .$$

Demostración. Considerando las propiedades de la esperanza de una matriz, obtenemos:

$$\begin{aligned} V(\mathbf{X}) &= E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'] = \\ &= E[\mathbf{X}\mathbf{X}' - \mathbf{X}(E(\mathbf{X}))' - E(\mathbf{X})\mathbf{X}' + E(\mathbf{X})(E(\mathbf{X}))'] = \\ &= E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})(E(\mathbf{X}))' , \end{aligned}$$

sin más que tener en cuenta que $E(\mathbf{X}') = (E(\mathbf{X}))'$. ♣

P2. Sea $\mathbf{X}$ una variable aleatoria n-dimensional y sea $\mathbf{C}$ una matriz $n \times 1$ de constantes. Entonces,

$$V(\mathbf{X} + \mathbf{C}) = V(\mathbf{X}) .$$

Demostración. Como $E(\mathbf{X} + \mathbf{C}) = E(\mathbf{X}) + \mathbf{C}$, entonces,

$$(\mathbf{X} + \mathbf{C}) - E(\mathbf{X} + \mathbf{C}) = \mathbf{X} - E(\mathbf{X}) .$$

En consecuencia, desarrollando la matriz de varianzas-covarianzas de $\mathbf{X} + \mathbf{C}$,

$$\begin{aligned} V(\mathbf{X} + \mathbf{C}) &= E[(\mathbf{X} + \mathbf{C} - E(\mathbf{X} + \mathbf{C})) \cdot (\mathbf{X} + \mathbf{C} - E(\mathbf{X} + \mathbf{C}))'] = \\ &= E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'] = V(\mathbf{X}) . \clubsuit \end{aligned}$$

P3. Sea $\mathbf{X}$ una variable aleatoria n-dimensional y sea $\mathbf{C}$ una matriz $m \times n$ de constantes. Entonces,

$$V(\mathbf{C} \cdot \mathbf{X}) = \mathbf{C} \cdot V(\mathbf{X}) \cdot \mathbf{C}' .$$

Demostración. Hallemos la matriz de varianzas-covarianzas de $\mathbf{C}\cdot\mathbf{X}$ y apliquemos las propiedades de la esperanza de matrices:

$$\begin{aligned} V(\mathbf{C}\cdot\mathbf{X}) &= E[(\mathbf{C}\cdot\mathbf{X}-E(\mathbf{C}\cdot\mathbf{X}))\cdot(\mathbf{C}\cdot\mathbf{X}-E(\mathbf{C}\cdot\mathbf{X}))'] = \\ &= E[\mathbf{C}(\mathbf{X}-E(\mathbf{X}))\cdot(\mathbf{X}-E(\mathbf{X}))'\mathbf{C}'] = \\ &= \mathbf{C}\cdot V(\mathbf{X})\cdot\mathbf{C}' \,. \clubsuit \end{aligned}$$

Las dos propiedades anteriores se resumen en el siguiente resultado que proporciona la matriz de varianzas-covarianzas de una variable aleatoria m-dimensional $\mathbf{Y}$, cuyas componentes son combinaciones lineales de las componentes de $\mathbf{X}$. Así, si

$$\mathbf{Y} = \mathbf{A}\cdot\mathbf{X}+\mathbf{B}\,,$$

donde $\mathbf{A}$ y $\mathbf{B}$ son dos matrices de constantes de dimensión $m\times n$ y $m\times 1$, respectivamente, y $\mathbf{X}$ es una variable aleatoria n-dimensional, entonces,

$$V(\mathbf{Y}) = V(\mathbf{A}\cdot\mathbf{X}+\mathbf{B}) = \mathbf{A}\cdot V(\mathbf{X})\cdot\mathbf{A}'\,.$$

P4. La matriz de varianzas-covarianzas, $V(\mathbf{X})$, es una matriz simétrica.

Demostración. La demostración es inmediata, ya que $Cov(X_i,X_j) = Cov(X_j,X_i)$ y, por tanto, $\sigma_{ij}=\sigma_{ji}$. ♣

P5. La matriz de varianzas-covarianzas, $V(\mathbf{X})$, es una matriz semidefinida positiva.

Demostración. Sea $\mathbf{B}=(b_1,\ldots,b_n)$ una matriz $1\times n$ cualquiera de números reales; entonces,

$$\mathbf{B}\cdot V(\mathbf{X})\cdot\mathbf{B}' = V(\mathbf{B}\cdot\mathbf{X}) = Var(b_1\cdot X_1+\cdots+b_n\cdot X_n)\geq 0\,,$$

ya que $\mathbf{B}\cdot\mathbf{X}$ es una variable aleatoria unidimensional y, por consiguiente, su matriz de varianzas-covarianzas está formada por un único elemento: la varianza de $\mathbf{B}\cdot\mathbf{X}$.

Luego, $V(\mathbf{X})$ es una matriz semidefinida positiva. ♣

P6. La matriz de varianzas-covarianzas, $V(\mathbf{X})$, es una matriz definida positiva si, y sólo si, no existe ninguna combinación lineal entre las variables $X_1,\ldots,X_n$.

Demostración. A la vista de la demostración anterior, para todo $\mathbf{B}=(b_1,\ldots,b_n)$ no idénticamente nulo, la condición

$$\mathbf{B}\cdot V(\mathbf{X})\cdot\mathbf{B}'>0$$

es equivalente a

$$Var(b_1\cdot X_1+\cdots+b_n\cdot X_n)>0\,.$$

Pero esta última condición equivale a que la variable aleatoria

$$b_1\cdot X_1+\cdots+b_n\cdot X_n$$

no sea constante. En definitiva,

$$\mathbf{B} \cdot V(\mathbf{X}) \cdot \mathbf{B}' > 0$$

si, y sólo si,

$$b_1 \cdot X_1 + \cdots + b_n \cdot X_n \neq c \,,$$

es decir, si, y sólo si, no existe una combinación lineal entre las variables. ♣

Ejemplo 7.11. Sea (X, Y) una variable aleatoria bidimensional continua cuya función de densidad conjunta ya se estudió en anteriores ejemplos,

$$f_{(X,Y)}(x,y) = \begin{cases} e^{-x} & 0 < y < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

Vamos a calcular la matriz de esperanzas y la matriz de varianzas-covarianzas de (X, Y). Como vimos en el ejemplo 7.8, los principales momentos de (X, Y) valen:

$$E(X) = 2\,, \quad E(Y) = 1\,, \quad Var(X) = 2\,, \quad Var(Y) = 1 \quad \text{y} \quad Cov(X,Y) = 1$$

Entonces, la matriz de esperanzas será

$$\begin{pmatrix} E(X) \\ E(Y) \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

y la matriz de varianzas-covarianzas,

$$\begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Si consideramos nuevas variables, combinaciones lineales de X y de Y, podemos hallar con facilidad sus matrices de esperanzas y de varianzas-covarianzas utilizando las propiedades demostradas en este mismo apartado. Así, las variables

$$U = 2X - 3Y + 4$$

y

$$V = X + 5Y$$

se pueden expresar matricialmente como

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

De aquí,

$$\begin{pmatrix} E(U) \\ E(V) \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} E(X) \\ E(Y) \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \end{pmatrix},$$

y la matriz de varianzas-covarianzas resulta del producto

$$\begin{pmatrix} 2 & -3 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 5 \end{pmatrix},$$

con lo cual

$$\begin{pmatrix} Var(U) & Cov(U,V) \\ Cov(U,V) & Var(V) \end{pmatrix} = \begin{pmatrix} 5 & -4 \\ -4 & 37 \end{pmatrix}.$$

Puede comprobar el lector que esta última matriz es definida positiva; en cambio, si consideramos una nueva variable,

$$W = X - 21Y + 8\,,$$

la correspondiente matriz de varianzas-covarianzas,

$$\begin{pmatrix} Var(U) & Cov(U,V) & Cov(U,W) \\ Cov(U,V) & Var(V) & Cov(V,W) \\ Cov(U,W) & Cov(V,W) & Var(W) \end{pmatrix} = \begin{pmatrix} 5 & -4 & 22 \\ -4 & 37 & -119 \\ 22 & -119 & 401 \end{pmatrix},$$

es semidefinida positiva, puesto que existe una combinación lineal entre U, V y W; en concreto,

$$W = 2U - 3V\,. \triangleleft$$

7.6 Ejercicios propuestos

7.1. Sea (X,Y) una variable aleatoria bidimensional discreta y sea $Z = g(X,Y)$ una función de (X,Y). Demuéstrese que, si existen,

$$E[g(X,Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} g(x_i,y_j)\cdot p[X=x_i, Y=y_j]$$

y

$$E(Z) = \sum_{h=1}^{\infty} z_h \cdot p[Z=z_h]$$

coinciden.

7.2. Sea (X,Y) una variable aleatoria bidimensional continua y sea $Z = g(X,Y)$ una función de (X,Y), donde g verifica las condiciones de la proposición 5.28. Demuéstrese que, si existen,

$$E[g(X,Y)] = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} g(x,y)\cdot f_{(X,Y)}(x,y)dxdy$$

y

$$E(Z) = \int_{z=-\infty}^{\infty} z\cdot f_Z(z)dz$$

coinciden.

7.3. Sean X e Y dos variables aleatorias, siendo Y degenerada, es decir, constante. Pruébese que $Cov(X,Y) = 0$.

7.4. Sean $X_1,\ldots,X_n$ variables aleatorias incorrelacionadas de media μ y varianza σ^2, y sean $a_1,\ldots,a_n$ números reales cualesquiera.

a) Calcúlese la varianza de la combinación lineal $\sum_{i=1}^{n} a_iX_i$.

b) ¿Para qué valores $a_1,\ldots,a_n$ la combinación lineal anterior tiene media μ y varianza mínima?

7.5. Sean $X_1, \ldots, X_n$ variables aleatorias incorrelacionadas de varianza σ^2, y sean $a_1, \ldots, a_n$ y $b_1, \ldots, b_n$ constantes. ¿Qué deben cumplir estas constantes para que $\sum_{i=1}^n a_i X_i$ y $\sum_{i=1}^n b_i X_i$ estén incorrelacionadas?

7.6. Sean X, Y y Z variables aleatorias incorrelacionadas con igual varianza. Hállese el coeficiente de correlación entre $X + Y$ y $X + Z$.

7.7. Sean X e Y los salarios mensuales percibidos por dos empleados de un banco. Suponiendo que estas dos variables están incorrelacionadas:

a) Hállese el coeficiente de correlación entre la ganancia conjunta de estos empleados y su diferencia salarial.

b) Demuéstrese que si X e Y están igualmente distribuidas, las variables definidas en el apartado anterior están incorrelacionadas.

7.8. Sean X e Y dos variables aleatorias con varianza σ^2 y coeficiente de correlación ρ. Calcúlese la varianza de la media aritmética de X e Y. ¿Qué valor de ρ minimiza dicha varianza?

7.9. Sea X una variable aleatoria que toma los valores -1, 0 y 1, todos ellos con la misma probabilidad. Demuéstrese que X y X^2 están incorrelacionadas, pero son dependientes.

7.10. Sean X e Y dos variables aleatorias independientes e igualmente distribuidas, cuya función de densidad es

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

Hállese el coeficiente de correlación entre las variables $X \cdot Y$ y $(1 - X) \cdot (1 - Y)$.

7.11. Se realiza un estudio sobre la condición de los trabajadores de un sector del pequeño comercio. Siendo X el número de trabajadores por establecimiento e Y el número de los que no pertenecen a la familia propietaria del mismo, se ha obtenido la siguiente función de probabilidad conjunta:

$X \backslash Y$	0	1	2
1	2/9	1/9	0
2	1/9	3/9	0
3	0	1/9	2/9

a) Hállense las curvas y rectas de regresión y estúdiese la bondad de los ajustes. Dibújense estas curvas y rectas sobre la correspondiente nube de puntos en un sistema cartesiano.

b) Calcúlese la función de distribución del número de trabajadores que pertenecen a la familia propietaria.

c) Si un establecimiento cuenta a lo sumo con dos trabajadores, ¿cuál será el número esperado de los que pertenecen a la familia?

7.12. Un estudio sociológico ha estimado que el número total de habitaciones, X, y el número de dormitorios, Y, de las viviendas de una gran ciudad, se distribuyen de acuerdo con la siguiente función de probabilidad:

$X\backslash Y$	1	2	3	4
1	5/100	0	0	0
2	10/100	0	0	0
3	0	35/100	0	0
4	0	10/100	35/100	0
5	0	0	1/100	2/100
6	0	0	0	2/100

a) Hállense las curvas y rectas de regresión y estúdiese la bondad de los ajustes.

b) Una vivienda tiene más de 3 habitaciones, ¿cuál es el número esperado de dormitorios?

7.13. Se elige un número al azar, X, entre los cuatro primeros números naturales: 1, 2, 3 y 4. Elegido un determinado número, se hace una segunda elección, Y, entre los mismos números, excluidos los inferiores al ya elegido.

a) Hállese la distribución de (X, Y).

b) Calcúlense las curvas y rectas de regresión, estudiando la bondad de los ajustes.

c) Dibújense sobre un sistema de ejes cartesianos la nube de puntos y las curvas y rectas de regresión.

d) Sabiendo que el segundo número elegido fue mayor que 2, hállese el valor esperado del producto de los dos números elegidos.

7.14. La variable aleatoria bidimensional (X, Y) representa la renta disponible y el ahorro asociado que, efectuando una cambio de unidades, tiene la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Calcúlese la recta de regresión que explica el ahorro en función de la renta.

b) Obténgase el ahorro esperado para un nivel de renta fijo x.

7.15. Con objeto de llevar a cabo una campaña publicitaria de lanzamiento de una nueva marca de cereales para el desayuno, se realiza un estudio estadístico sobre los hábitos en el desayuno de las familias de una cierta ciudad. De este análisis se obtiene que la proporción de individuos que toman leche en una familia, X, y la de los que toman cualquier tipo de infusión, Y, responde a la siguiente función de densidad conjunta:

$$f(x,y) = \begin{cases} 24xy & x \geq 0 \ , y \geq 0 \ , x+y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállense las rectas y curvas de regresión.

b) Estúdiese la bondad de los ajustes anteriores.

c) ¿Cuál es la proporción esperada de consumidores de leche en una familia en la que el 70 % de sus miembros toma infusiones?

d) Si en una familia la mitad de sus individuos toma leche, ¿cuál es la proporción esperada de individuos que no toman ni infusiones ni leche?

7.16. Una persona contrae el virus de una cierta enfermedad que necesita hospitalización. Sea X la variable aleatoria que designa el tiempo, en semanas, que tarda en manifestarse la enfermedad. Sea Y la variable aleatoria que expresa el tiempo total, también en semanas, desde que el virus es contraído hasta que el paciente es dado de alta. Se supone que en el mismo momento en que los síntomas aparecen el enfermo es hospitalizado.
Se ha estimado que la variable aleatoria (X, Y) tiene la siguiente función de densidad:

$$f(x,y) = \begin{cases} xe^{-y} & \text{si } 0 \leq x \leq y < \infty \\ 0 & \text{en el resto} \end{cases}$$

a) Hállense las curvas y las rectas de regresión y estúdiese la bondad de los ajustes.

b) Represéntense las curvas y rectas de regresión en un sistema de ejes cartesianos sobre el recinto de valores posibles de (X, Y).

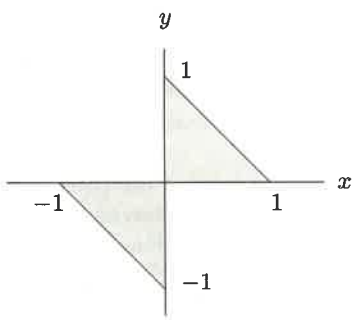
c) Calcúlese la probabilidad de que un paciente esté hospitalizado durante menos de una semana.

d) Suponiendo que un paciente sale hoy del hospital después de 5 días de tratamiento, ¿cuál es la probabilidad de que contrajera el virus hace menos de dos semanas?

7.17. Sea (X, Y) una variable aleatoria bidimensional cuya función de densidad conjunta es

$$f(x,y) = \begin{cases} \frac{1}{2} & \text{si } (x,y) \in S \\ 0 & \text{en el resto} \end{cases}$$

donde S es el recinto sombreado en la siguiente figura:



a) Compruébese que las variables son dependientes, pero están incorrelacionadas.

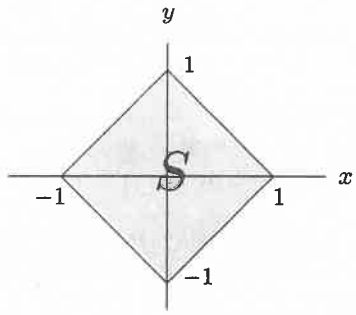
b) Hállense las curvas y rectas de regresión y discútase la bondad de los ajustes.

c) Calcúlese $p\left[X \cdot Y > \frac{1}{8}\right]$.

7.18. Sea (X, Y) una variable aleatoria bidimensional con función de densidad

$$f(x,y) = \begin{cases} k & \text{si } (x,y) \in S \\ 0 & \text{en el resto} \end{cases}$$

donde S es el recinto señalado en la siguiente figura:



a) Hállese el valor de la constante k.

b) Estúdiese la independencia y la incorrelación entre X e Y.

c) Obténganse las rectas y las curvas de regresión y coméntese el ajuste en todos los casos.

d) Calcúlese el valor esperado de Y, si $X - Y$ es igual a $\frac{1}{4}$.

7.19. Se quiere estudiar la renta mensual, X, y el consumo en alimentación, Y, de las familias de un cierto país. El consumo mensual en alimentación, en cientos de miles de pesetas, de las familias con una cantidad x de renta, se distribuye según la siguiente función de densidad:

$$f_{(Y|X=x)}(y) = \begin{cases} 8\frac{y}{x^2} & 0 < y < \frac{x}{2} \\ 0 & \text{en el resto} \end{cases}$$

Por su parte, la renta mensual, X, expresada en las mismas unidades, es una variable aleatoria cuya función de densidad es:

$$f_X(x) = \begin{cases} \frac{1}{2}x^2 e^{-x} & 0 < x < \infty \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese la mejor explicación lineal del consumo en alimentación a partir de la renta mensual. Estúdiese la bondad de este ajuste.

b) Si una familia tiene unos ingresos mensuales de 250 mil pesetas, ¿cuál es la cantidad esperada que dedica a alimentación?

c) Obténgase la renta mensual media de una familia, sabiendo que gasta 50 mil pesetas en alimentación.

7.20. Dada una variable aleatoria (X, Y), interesa explicar la variable Y a partir de la variable X, bien mediante una curva del tipo $y = a \cdot |x|$, bien mediante una parábola centrada en el origen, $y = b \cdot x^2$.

a) ¿Cuáles son los valores de a y de b que nos proporcionan un mejor ajuste, en cada caso, por el método de mínimos cuadrados?

b) Hállense las funciones de regresión obtenidas en el apartado anterior para la variable aleatoria bidimensional (X, Y) cuya distribución viene dada por la tabla:

$X\backslash Y$	0	1	2
-1	0	1/6	1/6
0	2/6	0	0
1	0	2/6	0

c) De acuerdo con el criterio mínimo cuadrático, ¿cuál de las dos regresiones explica mejor Y a partir de X en la distribución anterior?

d) Calcúlese $E(Y \mid X = x)$ y $Var(Y \mid X = x)$.

7.21. Sea (X, Y) una variable aleatoria bidimensional. Sea $\tilde{Y}$ la función de X dada por la recta de regresión de Y sobre X, y ε el error de la regresión.

a) Compruébese que $\rho_{Y,\tilde{Y}} = \rho_{X,Y}$.

b) Demuéstrese que $\tilde{Y}$ y ε están incorrelacionadas.

c) Hállese el coeficiente de correlación entre X e $\tilde{Y}$ y entre X y ε.

d) Demuéstrese que $\rho^2_{Y,\varepsilon} = 1 - \rho^2_{X,Y}$.

e) Sea (X, Y) una variable aleatoria cuya función de probabilidad es:

$X\backslash Y$	-1	0	1
-1	2/6	0	1/6
1	1/6	1/6	1/6

Compruébense, para esta distribución, todos los resultados anteriores.

7.22. Sean X, Y y Z tres variables de esperanza cero. Se define $\rho_{XY \mid Z}$ como el coeficiente de correlación entre las variables $X - b_{X \mid Z} \cdot Z$ e $Y - b_{Y \mid Z} \cdot Z$ donde

$$b_{X \mid Z} = \frac{Cov(X, Z)}{Var(Z)} \qquad \text{y} \qquad b_{Y \mid Z} = \frac{Cov(Y, Z)}{Var(Z)} ,$$

es decir, los coeficientes de la regresión de X sobre Z y de Y sobre Z, respectivamente. Demuéstrese que

$$\rho_{XY \mid Z} = \frac{\rho_{XY} - \rho_{XZ} \cdot \rho_{YZ}}{(1 - \rho^2_{XZ})^{1/2} \cdot (1 - \rho^2_{YZ})^{1/2}} .$$

7.23. Sea (X, Y) una variable aleatoria bidimensional cuya función de densidad es

$$f(x, y) = \begin{cases} \frac{1}{x} & 0 < y < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Calcúlese $E(X \mid Y = y)$, $Var(X \mid Y = y)$, $E(Y \mid X = x)$ y $Var(Y \mid X = x)$.

b) Hállese la matriz de esperanzas y la matriz de varianzas-covarianzas de (X, Y).

c) Obténganse las esperanzas, varianzas y covarianza de U y V, donde $U = 2X + 3Y$ y $V = X - 2Y$.

7.24. Sean $Y_1, \ldots, Y_n$ variables aleatorias incorrelacionadas con media 0 y varianza σ^2. Definimos las variables $X_1 = Y_1$ y $X_i = \alpha \cdot X_{i-1} + Y_i$ $(i = 2, \ldots, n)$, donde α es una constante. Hállese la matriz de esperanzas y la matriz de varianzas-covarianzas de $(X_1, \ldots, X_n)$.

Capítulo 8

FUNCIONES CARACTERÍSTICAS

En los dos capítulos anteriores hemos estudiado las características más importantes de las variables aleatorias, que resumen aspectos concretos de sus distribuciones de probabilidad. No obstante, en general, éstas no caracterizan a las variables en el sentido de que, conocidas dichas características, quede determinada la función de distribución, o la función de probabilidad o de densidad, según el caso.

Ciertas excepciones a dicha regla general son el resultado de la información disponible sobre el tipo de variable de que se trata. Por ejemplo, supongamos que X es una variable aleatoria que vale la unidad si se obtiene cara al lanzar una moneda, y cero si se obtiene cruz; su función de probabilidad es $p[X = 1] = p$ y $p[X = 0] = 1 - p$, donde p es la probabilidad de obtener cara. Además, p es la esperanza de X, por lo que, en este caso, el conocimiento de la esperanza determina la función de probabilidad de la variable.

La función característica de una variable aleatoria es una función asociada a su distribución de probabilidades que permite conocer esta última. En este sentido, constituye una poderosa herramienta en la investigación de las distribuciones de probabilidad de variables aleatorias de una o varias dimensiones, empleada especialmente en el estudio de las sumas de variables independientes. De hecho, a partir de que fueron estudiadas sus propiedades por P. LÉVY en el tercer decenio de nuestro siglo, se han resuelto problemas importantes que estaban abiertos desde doscientos años antes como, por ejemplo, los relacionados con el teorema del límite central que veremos en el capítulo 13.

Desgraciadamente, su utilidad se ve empañada por su dificultad de cálculo, ya que la función característica de una variable aleatoria es la esperanza de una función compleja de la misma. Al nivel fijado para este libro, sólo podemos obtener con rigor funciones características de distribuciones de probabilidad muy sencillas; en la mayoría de los casos, utilizaremos sus propiedades para obtener la función característica de unas variables a partir de las de otras, suministrando al lector las funciones características de las variables básicas sin calcularlas.

Antes de definir la función característica de una variable aleatoria, es necesario estudiar ciertos conceptos.

8.1 Variable aleatoria compleja. Definición y esperanza matemática

Definición 8.1 *Sea (Ω, α, p) un espacio probabilístico y sean X e Y variables aleatorias (reales) sobre Ω. Llamamos variable aleatoria compleja a la aplicación del espacio muestral en el conjunto de los números complejos,*

$$Z : \Omega \longrightarrow \mathbb{C} ,$$

definida como

$$Z(\omega) = X(\omega) + i \cdot Y(\omega)$$

para todo elemento ω de Ω, siendo i la unidad imaginaria de los números complejos.

Esta variable se describe como $Z = X + i \cdot Y$.

OBSERVACIONES

O1. Recordemos que cualquier número complejo, z, se puede escribir en la denominada *forma binómica*, $z = a + i \cdot b$, donde a y b son números reales. El primero de ellos, a, se denomina *parte real* de z, y el segundo, b, *parte imaginaria*. Se puede, por tanto, representar dicho número complejo como el punto del plano (a, b), reservando el eje de abscisas para su parte real, y el de ordenadas para la parte imaginaria.

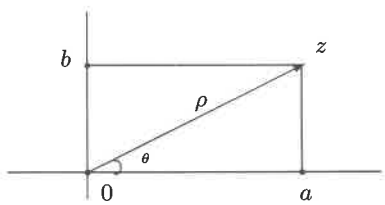


FIG. 8.1.

Los números complejos pueden describirse también en la *forma polar*, que caracteriza a uno cualquiera de ellos mediante su *módulo* y su *argumento*. Dado un complejo $z = a + ib$, su módulo, $|z|$, que designaremos comúnmente por ρ, es la distancia del punto (a, b) al origen, es decir, $\rho = \sqrt{a^2 + b^2}$; por su parte, su argumento, θ, es el ángulo que el vector de extremo inicial en el origen y extremo final en el punto (a, b) forma con el semieje positivo de abscisas. Se cumplirá, así, que $\operatorname{tg}(\theta) = \frac{b}{a}$ y, en consecuencia, que $\theta = \operatorname{arctg}(\frac{b}{a})$, para $a \neq 0$; cuando a se anula, el argumento θ valdrá $\frac{\pi}{2}$, si b es positivo, y $\frac{3\pi}{2}$ si b es negativo. La forma polar del número complejo sería, entonces, ρ_θ.

Las partes real e imaginaria de un número complejo pueden escribirse en función de las coordenadas polares; concretamente, $a = \rho \cdot \cos(\theta)$ y $b = \rho \cdot \operatorname{sen}(\theta)$. En definitiva, el número complejo $z = a + ib$ puede expresarse como

$$z = a + i \cdot b = \rho \cdot \cos(\theta) + i \cdot \rho \cdot \operatorname{sen}(\theta) .$$

La figura 8.1 ilustra las relaciones entre ambas caracterizaciones de un número complejo.

Una tercera forma de describir los números complejos resultará a veces conveniente: la *forma exponencial.* Dado un ángulo θ, se define la exponencial compleja $e^{i\theta}$ como

$$e^{i\theta} = \cos(\theta) + i \cdot \operatorname{sen}(\theta).$$

En consecuencia,

$$\begin{aligned} z &= a + i \cdot b = \rho \cdot \cos(\theta) + i \cdot \rho \cdot \operatorname{sen}(\theta) = \\ &= \rho \cdot [\cos(\theta) + i \cdot \operatorname{sen}(\theta)] = \rho e^{i\theta} . \end{aligned}$$

Esta última expresión recibe el nombre de forma exponencial del complejo z.

Para los desarrollos posteriores será conveniente recordar que, dado un número complejo $z = a + ib$, su conjugado, $\bar{z}$, es el número complejo que tiene su misma parte real y opuesta parte imaginaria, esto es, $\overline{z} = a - ib$. Recordemos, asimismo, que los números complejos se multiplican en forma binómica con las reglas del producto de números reales, sin más que tener en cuenta que $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$ y así sucesivamente.

O2. Recordemos que si f es una aplicación compleja de un conjunto cualquiera, M, en $\mathbb{C}$,

$$f : M \longrightarrow \mathbb{C},$$

podemos escribir f en la forma $f(x) = g(x) + i \cdot h(x)$ para todo $x \in M$, donde g y h son aplicaciones reales que se denominan componente real y componente imaginaria, respectivamente, de la función f.

La variables aleatorias complejas son, por tanto, aplicaciones complejas del espacio muestral, Ω, en $\mathbb{C}$, que envían cada uno de sus elementos, ω, a un número complejo, $X(\omega) + i \cdot Y(\omega)$, donde X e Y son, en este caso, las componentes real e imaginaria, respectivamente.

O3. Es evidente que no todas las aplicaciones de Ω en $\mathbb{C}$ son variables aleatorias complejas. La definición exige que la parte real, X, y la parte imaginaria, Y, en las cuales cualquier aplicación compleja Z puede descomponerse, sean, a su vez, variables aleatorias.

O4. Las variables aleatorias definidas en el capítulo 4 pueden considerarse, según la definición anterior, como un caso particular de las variables aleatorias complejas, en las que la parte imaginaria es igual a cero. Sin embargo, el objetivo de la definición 8.1 no es generalizar el concepto de variable aleatoria, sino sentar las bases para la definición de función característica.

La siguiente definición precisa cómo debe entenderse la esperanza de una variable aleatoria compleja.

Definición 8.2 *Sea (Ω, α, p) un espacio probabilístico y sea $Z = X + i \cdot Y$ una variable aleatoria compleja. Se define la esperanza matemática de Z, siempre que exista, como*

$$E(Z) = E(X) + i \cdot E(Y) .$$

Como indica la definición, la esperanza matemática de una variable aleatoria compleja es un número complejo que existe siempre que existan su parte real y su parte imaginaria, es decir, que existan las esperanzas de X e Y, de acuerdo con la condición que impusimos en el capítulo 6.

Es posible definir otros momentos y características de una variable aleatoria compleja; sin embargo, nuestro interés no está en el estudio exhaustivo de las variables aleatorias complejas, sino sólo en la presentación de aquellos conceptos necesarios para definir la función característica.

8.2 Función característica de una variable aleatoria unidimensional. Propiedades

Una vez introducidos los conceptos previos necesarios, pasemos a definir el objeto de estudio de este capítulo, esto es, lo que entendemos por una función característica.

Definición 8.3 *Sea (Ω, α, p) un espacio probabilístico y sea X una variable aleatoria (real) sobre Ω. Llamamos función característica de X a la aplicación*

$$\varphi_X : \mathbb{R} \longrightarrow \mathbb{C} ,$$

donde para todo número real, t,

$$\varphi_X(t) = E(e^{itX}) .$$

Teniendo en cuenta la relación entre la forma exponencial y la forma binómica de los números complejos, podemos escribir

$$e^{itX} = \cos(tX) + i \cdot \operatorname{sen}(tX);$$

además, como las funciones seno y coseno son continuas, $\cos(tX)$ y $\operatorname{sen}(tX)$ son variables aleatorias reales y, por tanto, e^{itX} es una variable aleatoria compleja.

De la definición 8.3 se deduce que la función característica de una variable aleatoria X es la esperanza de una variable aleatoria compleja; en concreto,

$$\varphi_X(t) = E(e^{itX}) = E[\cos(tX)] + i \cdot E[\operatorname{sen}(tX)] .$$

Si X es una variable aleatoria discreta, se calculará como

$$\varphi_X(t) = \sum_{j=1}^{\infty} \cos(tx_j) p[X = x_j] + i \cdot \sum_{j=1}^{\infty} \operatorname{sen}(tx_j) p[X = x_j] =$$

$$= \sum_{j=1}^{\infty} [\cos(tx_j) + i \cdot \operatorname{sen}(tx_j)] \cdot p[X = x_j] =$$

$$= \sum_{j=1}^{\infty} e^{itx_j} \cdot p[X = x_j] .$$

Observe el lector que la serie anterior lo es en el campo complejo, cuyo estudio pormenorizado sobrepasa los límites fijados para este libro.

Si X es continua, el cálculo puede realizarse como

$$\varphi_X(t) = \int_{-\infty}^{\infty} \cos(tx) f_{X(x)}\, dx + i \cdot \int_{-\infty}^{\infty} \operatorname{sen}(tx) f_{X(x)}\, dx =$$

$$= \int_{-\infty}^{\infty} [\cos(tx) + i \cdot \operatorname{sen}(tx)] f_{X(x)}\, dx =$$

$$= \int_{-\infty}^{\infty} e^{itx} f_{X(x)}\, dx .$$

Asimismo, esta última integral es una integral en el campo complejo, encontrándonos ante las mismas limitaciones señaladas para el caso discreto.

Aunque, por lo general, el cálculo de funciones características rebasa los límites de este libro, plantearemos dos ejemplos con variables aleatorias sencillas donde dicho cálculo no presenta problemas.

Ejemplo 8.1. Consideremos la variable aleatoria X, que toma los valores 1 y 0, con probabilidades $\frac{1}{3}$ y $\frac{2}{3}$, respectivamente.

Como

$$e^{itX} = \begin{cases} e^{it} & \text{si } x = 1 \\ 1 & \text{si } x = 0 \end{cases}$$

la función característica de la variable aleatoria X es:

$$\varphi_X(t) = E(e^{itX}) = e^{it} \cdot p[X = 1] + 1 \cdot p[X = 0] =$$

$$= \frac{1}{3} e^{it} + \frac{2}{3} . \triangleleft$$

Ejemplo 8.2. Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

Como X es una variable aleatoria continua, su función característica será:

$$\varphi_X(t) = \int_0^1 e^{itX} \cdot 1\, dx = \int_0^1 \cos(tx)\, dx + i \cdot \int_0^1 \operatorname{sen}(tx)\, dx ,$$

que, cuando t sea distinta de cero, valdrá

$$\varphi_X(t) = \frac{\operatorname{sen}(t)}{t} + i \cdot \frac{[1 - \cos(t)]}{t} = \frac{\operatorname{sen}(t) + i - i \cdot \cos(t)}{t} .$$

Si multiplicamos y dividimos por i a la expresión anterior, recordando que $i^2 = -1$,

$$\varphi_X(t) = \frac{i \cdot \operatorname{sen}(t) + i^2 - i^2 \cdot \cos(t)}{it} = \frac{\cos(t) + i \cdot \operatorname{sen}(t) - 1}{it} = \frac{e^{it} - 1}{it} .$$

Para t igual a cero,

$$\varphi_X(0) = \int_0^1 \cos(0 \cdot x)\, dx + i \cdot \int_0^1 \operatorname{sen}(0 \cdot x)\, dx = \int_0^1 1\, dx + i \cdot \int_0^1 0\, dx = 1 + 0 \cdot i = 1 .$$

En definitiva,

$$\varphi_X(t) = \begin{cases} \frac{e^{it}-1}{it} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

es la función característica de la variable X. ◁

La función característica es un tipo de función de uso común en el análisis matemático; se trata de la denominada transformada de Fourier de la función de probabilidad o de densidad. Existe una transformación con aspecto muy similar a la transformada de Fourier y que se conoce como transformada de Laplace. En el estudio de las distribuciones de probabilidad, esta función se concreta en la denominada **función generatriz de momentos**, que se define como *una aplicación*

$$\phi_X : \mathbb{R} \longrightarrow \mathbb{R}$$

tal que, para todo número real t,

$$\phi_X(t) = E(e^{tX}) .$$

Por tanto, si X es una variable aleatoria discreta, su cálculo se realiza como

$$\phi_X(t) = \sum_{j=1}^{\infty} e^{tx_j} \cdot p[X = x_j] ,$$

y si X es continua,

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x) dx .$$

Con la función generatriz de momentos se obtienen los mismos resultados que con la función característica y, además, presenta la ventaja de que sus valores se encuentran en la recta real, evitando la complicación que supone operar con series e integrales en el campo complejo. Pero la función generatriz de momentos tiene un inconveniente que hace que desestimemos su uso: no existe para todas las distribuciones (si bien es verdad que existe para aquellas con las que nosotros habitualmente trabajamos).

La siguiente proposición confirma que la función característica existe siempre, lo que la hace preferible a la función generatriz de momentos.

Proposición 8.1 *La función característica, φ_X, de una variable aleatoria, X, existe siempre.*

Demostración. Para asegurar la existencia de la esperanza de una variable compleja, hemos de probar la existencia de las esperanzas de su parte real y de su parte imaginaria. Como sabemos,

$$\varphi_X(t) = E(e^{itX}) = E[\cos(tX) + i \cdot \operatorname{sen}(tX)] = E[\cos^2(tX)] + E[\operatorname{sen}^2(tX)] .$$

Tanto $\cos(tX)$ como $\operatorname{sen}(tX)$ son variables aleatorias acotadas, en concreto entre -1 y 1, por lo cual existen sus esperanzas. De aquí deducimos que también existe φ_X, de la que aquéllas son su parte real e imaginaria, respectivamente. ♣

Demostremos, a continuación, algunos de los resultados que justifican, en gran parte, la introducción del concepto de función característica. En ellos se muestra cómo es posible obtener con cierta facilidad esta función para una variable que depende linealmente de otra u otras, siempre que sean conocidas las de éstas.

Proposición 8.2 *Sea X una variable aleatoria y φ_X su función característica. Entonces, para cualesquiera números reales a y b,*

$$\varphi_{aX+b}(t) = e^{itb} \cdot \varphi_X(at) .$$

Demostración. Dada la variable aleatoria $aX + b$, su función característica es, por definición,

$$\begin{aligned} \varphi_{aX+b}(t) &= E\left[e^{it(aX+b)}\right] = E\left(e^{itaX} \cdot e^{itb}\right) = \\ &= E[(\cos(taX) + i \cdot \operatorname{sen}(taX)) \cdot (\cos(tb) + i \cdot \operatorname{sen}(tb))] . \end{aligned}$$

Operando en la última expresión, y agrupando términos,

$$\begin{aligned} \varphi_{aX+b}(t) &= E[(\cos(taX)\cos(tb) - \operatorname{sen}(taX)\operatorname{sen}(tb)) + \\ &\quad + i(\cos(taX)\operatorname{sen}(tb) + \operatorname{sen}(taX)\cos(tb))] = \\ &= E[\cos(taX)\cos(tb) - \operatorname{sen}(taX)\operatorname{sen}(tb)] + \\ &\quad + iE[\cos(taX)\operatorname{sen}(tb) + \operatorname{sen}(taX)\cos(tb)] . \end{aligned}$$

Teniendo en cuenta que, $\cos(tb)$ y $\operatorname{sen}(tb)$ son constantes, la expresión anterior resulta:

$$\begin{aligned} \varphi_{aX+b}(t) &= \cos(tb)E[\cos(taX)] - \operatorname{sen}(tb)E[\operatorname{sen}(taX)] + \\ &\quad + i\operatorname{sen}(tb)E[\cos(taX)] + i\cos(tb)E[\operatorname{sen}(taX)] = \\ &= [\cos(tb) + i\operatorname{sen}(tb)] \cdot [E(\cos(taX)) + E(i\operatorname{sen}(taX))] = \\ &= e^{itb}E\left(e^{itaX}\right) = e^{itb} \cdot \varphi_X(at) . \clubsuit \end{aligned}$$

La demostración anterior resulta farragosa al tener que transformar las esperanzas de variables complejas en esperanzas reales. Si aceptamos que con las esperanzas de variables aleatorias complejas se opera como con las de las reales, se obtiene

el resultado de una manera más sencilla (adviértase que esta forma de proceder, si bien es razonable, no es rigurosa).

Como ya vimos,

$$\varphi_{aX+b}(t) = E\left[e^{it(aX+b)}\right] = E\left(e^{itaX} \cdot e^{itb}\right) .$$

Ahora bien, como e^{itb} es una constante, si aceptamos que la esperanza de una constante por una variable aleatoria compleja es el producto de la constante por la esperanza de la variable,

$$E\left(e^{itaX} \cdot e^{itb}\right) = e^{itb} \cdot E\left(e^{itaX}\right) = e^{itb} \cdot \varphi_X(at) .$$

Por tanto, obtendríamos el resultado,

$$\varphi_{aX+b}(t) = e^{itb} \cdot \varphi_X(at) .$$

De la proposición anterior se deducen los dos corolarios siguientes.

Corolario 8.2.1 *La función característica de la variable aleatoria aX vale*

$$\varphi_{aX}(t) = \varphi_X(at) .$$

Basta hacer b igual a cero en la proposición anterior.

Corolario 8.2.2 *La función característica de la variable aleatoria $X + b$ vale*

$$\varphi_{X+b}(t) = e^{itb} \cdot \varphi_X(t) .$$

Haciendo $a = 1$ en la proposición anterior, se concluye el resultado de inmediato.

La siguiente proposición permite obtener la función característica de una suma de variables aleatorias independientes a partir de las funciones características de cada una de las variables.

Proposición 8.3 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes y sean φ_{X_1}, ..., φ_{X_n}, sus respectivas funciones características. Entonces,*

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) .$$

Demostración. Realizaremos inducción matemática sobre n, demostrándolo primero para $n = 2$ y, suponiendo que se cumple para un valor concreto de n, $n = k$, lo comprobaremos para $k + 1$.

Para $n = 2$, consideraremos las variables independientes X_1 y X_2:

$$\begin{aligned} \varphi_{X_1+X_2}(t) &= E\left[e^{it(X_1+X_2)}\right] = E[e^{itX_1}e^{itX_2}] = \\ &= E[(\cos(tX_1) + i\operatorname{sen}(tX_1)) \cdot (\cos(tX_2) + i\operatorname{sen}(tX_2))] , \end{aligned}$$

expresión que, operando y aplicando la definición de esperanza de una variable compleja, es igual a

$$E[\cos(tX_1)\cos(tX_2) - \operatorname{sen}(tX_1)\operatorname{sen}(tX_2)] + \\ + iE[\cos(tX_1)\operatorname{sen}(tX_2) + \operatorname{sen}(tX_1)\cos(tX_2)] ,$$

que resulta,

$$E[\cos(tX_1)] \cdot E[\cos(tX_2)] - E[\operatorname{sen}(tX_1)] \cdot E[\operatorname{sen}(tX_2)] + \\ + i[E(\cos(tX_1)) \cdot E(\operatorname{sen}(tX_2)) + E(\operatorname{sen}(tX_1)) \cdot E(\cos(tX_2))] ,$$

sin más que tener en cuenta la independencia de X_1 y X_2 y, por tanto, la de cualquier función continua de X_1 y X_2 (en este caso senos y cosenos). Así, obtendremos:

$$[E(\cos(tX_1)) + iE(\operatorname{sen}(tX_1))] \cdot [E(\cos(tX_2)) + iE(\operatorname{sen}(tX_2))] = \\ = E(e^{itX_1}) \cdot E(e^{itX_2}) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t) .$$

Supongamos que el resultado es cierto para un $n = k$ fijo ($k \geq 2$),

$$\varphi_{\sum_{j=1}^{k} X_j}(t) = \prod_{j=1}^{k} \varphi_{X_j}(t) ,$$

y probémoslo para $k+1$. Así,

$$\varphi_{\sum_{j=1}^{k+1} X_j}(t) = E\left(e^{\sum_{j=1}^{k+1} itX_j}\right) = E\left(e^{\sum_{j=1}^{k} X_j} \cdot e^{itX_{k+1}}\right) = \\ = E\left(e^{\sum_{j=1}^{k} X_j}\right) \cdot E(e^{itX_{k+1}}) ,$$

ya que el resultado es cierto para $n = 2$; como también lo es para k, tendremos:

$$\varphi_{\sum_{j=1}^{k+1} X_j}(t) = \prod_{j=1}^{k} E(e^{itX_j}) \cdot E(e^{itX_{k+1}}) = \prod_{j=1}^{k+1} E(e^{itX_j}) = \prod_{j=1}^{k+1} \varphi_{X_j}(t) .$$

En definitiva, hemos comprobado el resultado de modo directo para $k = 2$ y demostrado que, si se cumple el enunciado para un valor k cualquiera, se cumple para el siguiente, $k+1$. El principio de inducción completa garantiza, entonces, que el enunciado se cumple para cualquier valor n. ♣

Puede probarse este resultado mediante un razonamiento que no es una demostración formal, ya que utilizaremos propiedades que, aunque razonables (y de hecho ciertas), no han sido demostradas; en concreto, se utiliza la idea de que la esperanza del producto de variables complejas independientes es el producto de sus esperanzas, y la de que si n variables aleatorias son independientes, funciones complejas de cada una de ellas son también independientes. Estos resultados están demostrados únicamente en el campo real.

Veamos dicha ilustración.

$$\varphi_{\sum_{j=1}^{n} X_j}(t) = E\left(e^{\sum_{j=1}^{n} itX_j}\right) = E(e^{itX_1} e^{itX_2} \cdots e^{itX_n}) .$$

Pero, si $X_1, X_2, \ldots, X_n$ son independientes, lo serán las variables complejas e^{itX_1}, e^{itX_2}, ..., e^{itX_n}. Por tanto,

$$\begin{aligned} \varphi_{\sum_{j=1}^n X_j}(t) &= E(e^{itX_1}e^{itX_2}\cdots e^{itX_n}) = \\ &= E(e^{itX_1})\cdot E(e^{itX_2})\cdots E(e^{itX_n}) = \\ &= \varphi_{X_1}(t)\cdot\varphi_{X_2}(t)\cdots\varphi_{X_n}(t)\,, \end{aligned}$$

lo que probaría el resultado.

De la proposición anterior se deducen dos importantes corolarios. Como dijimos en la introducción, la mayor utilidad de las funciones características reside en su facilidad para obtener la distribución de sumas de variables aleatorias. Veamos, de momento, cómo se obtiene la función característica de una suma de variables, si éstas tienen la misma distribución de probabilidades.

Corolario 8.3.1 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes e igualmente distribuidas. Entonces,*

$$\varphi_{\sum_{j=1}^n X_j}(t) = [\varphi_X(t)]^n\,,$$

donde φ_X es la función característica común a las n variables.

Demostración. Teniendo en cuenta la proposición 8.3, y dado que las funciones características de las variables serán iguales por estar idénticamente distribuidas,

$$\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) = [\varphi_X(t)]^n\,. \clubsuit$$

En particular, tenemos el resultado siguiente.

Corolario 8.3.2 *Sean n variables aleatorias, $X_1, \ldots, X_n$, independientes e igualmente distribuidas. Entonces, la función característica de la media muestral, $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$, será*

$$\varphi_{\overline{X}}(t) = \left[\varphi_X\left(\frac{t}{n}\right)\right]^n\,.$$

Es trivial utilizando los corolarios 8.2.1 y 8.3.1.

A partir de la función característica, podemos obtener momentos no centrales de la distribución, como veremos en el siguiente resultado. Su demostración exige, en el caso continuo, la derivación de una integral compleja bajo el signo integral. Esta derivación se realiza igual que para las integrales reales que dependen de un parámetro.

Obsérvese, por otro lado, que derivando e^{itx} con respecto a t, es decir, derivando la parte real y la parte imaginaria, obtenemos:

$$\begin{aligned}\frac{\partial}{\partial t}e^{itx} &= \frac{\partial}{\partial t}[\cos(tx) + i \cdot \operatorname{sen}(tx)] = -x \cdot \operatorname{sen}(tx) + ix \cdot \cos(tx) = \\ &= ix[\cos(tx) + i \cdot \operatorname{sen}(tx)] = ixe^{itx} .\end{aligned}$$

En otras palabras, las exponenciales complejas se derivan como las reales. Este hecho será utilizado en la demostración del siguiente resultado.

Proposición 8.4 *Sea φ_X la función característica de una variable aleatoria X. Si existe el momento no central de orden k, $(k = 1, 2, \ldots)$, α_k, existe la derivada de orden k en $t = 0$ de φ_X, y dicho momento vale*

$$\alpha_k = E(X^k) = \frac{1}{i^k}\frac{\partial^k}{\partial t^k}\varphi_X(t)\bigg|_{t=0} .$$

Demostración. La realizaremos en el caso continuo. Si X es discreta, la demostración es análoga.

$$\frac{\partial^k}{\partial t^k}\varphi_X(t) = \frac{\partial^k}{\partial t^k}E(e^{itX}) = \frac{\partial^k}{\partial t^k}\left[\int_{-\infty}^{\infty} e^{itx} \cdot f_X(x)\right] dx .$$

Si derivamos k veces bajo el signo integral, obtenemos:

$$\frac{\partial^k}{\partial t^k}\left[\int_{-\infty}^{\infty} e^{itx} \cdot f_X(x)dx\right] = \int_{-\infty}^{\infty}\frac{\partial^k}{\partial t^k}e^{itx} \cdot f_X(x)dx = i^k\int_{-\infty}^{\infty} x^k e^{itx} f_X(x)dx .$$

Pero, teniendo en cuenta que $|e^{itx}| = 1$,

$$\left|\int_{-\infty}^{\infty} x^k e^{itx} f_X(x)dx\right| \leq \int_{-\infty}^{\infty} |x^k| \cdot |e^{itx}| \cdot f_X(x)dx = \int_{-\infty}^{\infty} |x^k| \cdot f_X(x)dx < \infty ,$$

pues es el momento absoluto del orden k de X, que existirá al existir α_k. En consecuencia, también existe la integral que resulta al derivar k veces la función característica. Haciendo $t = 0$ en dicha integral,

$$\frac{\partial^k}{\partial t^k}\,\varphi_X(t)|_{t=0} = i^k\int_{-\infty}^{\infty} x^k e^{it0} f_X(x)dx = i^k\int_{-\infty}^{\infty} x^k f_X(x)dx = i^k\alpha_k .$$

Despejando, obtenemos el resultado deseado. ♣

La proposición anterior implica que la existencia del momento no central de orden k garantiza la derivabilidad de la función característica hasta dicho orden. El resultado recíproco, que garantizaría la existencia de momentos no centrales y, por tanto, de cualquier otro tipo (proposición 6.1), hasta un orden k siempre que la función característica de la variable aleatoria fuese derivable de orden k, sólo se cumple cuando k es un número par. Enunciemos el resultado, cuya demostración no incluimos ya que excede ampliamente el nivel y los objetivos propuestos para este libro. Este resultado garantiza, además, que si la función característica es derivable, su desarrollo en serie puede obtenerse en función de los momentos.

Proposición 8.5 *Sea φ_X la función característica de una variable aleatoria X. Si φ_X es derivable hasta el orden $2n$ en $t = 0$, existe el momento no central de orden k, α_k, para $k \leq 2n$.*

Además, si existe el momento no central de orden k, la función característica de X puede desarrollarse en serie de Mac Laurin de orden k,

$$\varphi_X(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!}\alpha_j + o(t^k) ,$$

para todo número real t, donde

$$\lim_{t\to 0} \frac{o(t^k)}{t^k} = 0 .$$

A continuación, veamos dos ejemplos que ilustran la proposición 8.4.

Ejemplo 8.3. Considemos la variable discreta del ejemplo 8.1:

$$X : \left| \begin{array}{ll} 0 & p[X=0] = \frac{2}{3} \\ 1 & p[X=1] = \frac{1}{3} \end{array} \right.$$

Su función característica, como ya calculamos, es

$$\varphi_X(t) = \frac{1}{3}e^{it} + \frac{2}{3} .$$

Derivando podremos hallar la esperanza:

$$\frac{\partial \varphi_X(t)}{\partial t} = \frac{1}{3} i e^{it} .$$

De aquí,

$$E(X) = \frac{1}{i} \cdot \left.\frac{\partial \varphi_X(t)}{\partial t}\right|_{t=0} = \frac{1}{i} \cdot i\frac{1}{3} = \frac{1}{3} .$$

Igualmente, calcularíamos el momento no central de orden 2:

$$E(X^2) = \frac{1}{i^2} \cdot \left.\frac{\partial^2 \varphi_X(t)}{\partial t^2}\right|_{t=0} = \frac{1}{i^2} \cdot i^2\frac{1}{3} = \frac{1}{3} ,$$

a partir del cual obtendremos la varianza,

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \left(\frac{1}{3}\right)^2 = \frac{2}{9} . \triangleleft$$

Ejemplo 8.4. En la variable aleatoria continua del ejemplo 8.2, cuya función de densidad es

$$f(x) = \left\{ \begin{array}{ll} 1 & 0 < x < 1 \\ 0 & \text{en el resto} \end{array} \right.$$

obtuvimos la función característica:

$$\varphi_X(t) = \begin{cases} \frac{e^{it}-1}{it} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

En este caso, resulta bastante más complicado derivar en $t = 0$, ya que hemos de aplicar la definición de derivada. De esta forma,

$$\varphi'_X(0) = \left.\frac{\partial \varphi_X(t)}{\partial t}\right|_{t=0} = \lim_{h\to 0} \frac{\varphi_X(h) - \varphi_X(0)}{h} = \lim_{h\to 0} \frac{\frac{e^{ih}-1}{ih} - 1}{h} = \lim_{h\to 0} \frac{e^{ih} - 1 - ih}{ih^2} .$$

El límite anterior es un caso de indeterminación, ya que tanto el numerador como el denominador tienden a cero. Si aplicamos la regla de L'Hôpital (derivando el numerador y el denominador respecto de h),

$$\lim_{h\to 0} \frac{e^{ih} - 1 - ih}{ih^2} = \lim_{h\to 0} \frac{ie^{ih} - i}{2ih} = \lim_{h\to 0} \frac{e^{ih} - i}{2h} .$$

Este límite vuelve a ser un caso de indeterminación del mismo tipo, por lo que volvemos a aplicar la misma regla y, en definitiva,

$$\varphi'_X(0) = \left.\frac{\partial \varphi_X(t)}{\partial}\right|_{t=0} = \lim_{h\to 0} \frac{e^{ih} - 1}{2h} = \lim_{h\to 0} \frac{ie^{ih}}{2} = \frac{i}{2} .$$

Por tanto,

$$E(X) = \frac{\varphi'_X(0)}{i} = \frac{1}{2} . \triangleleft$$

Compare el lector la dificultad (pequeña en el primer ejemplo, grande en el segundo) del cálculo de la esperanza a partir de la función característica, con la sencillez del cálculo directo de la misma. Sin embargo, hay casos en los que interesará obtener los momentos de esta forma.

La importancia de la función característica en el estudio de las distribuciones de variables aleatorias reside en el hecho de que caracteriza su distribución de probabilidades, como apuntábamos en la introducción de este capítulo. La siguiente proposición, que no demostraremos, muestra que la función de distribución de una variable aleatoria puede obtenerse a partir de la función característica.

No se utiliza el resultado para calcular funciones de distribución, sino para asignarlas a variables aleatorias cuya función característica sea conocida; expliquemos esta idea. Supongamos que se conoce la función de distribución, F_{X_1}, y la función característica, φ_{X_1}, de una variable aleatoria X_1, y que, para otra variable, X_2, se conoce únicamente su función característica, φ_{X_2}. Si ambas funciones características son iguales, esto es, si $\varphi_{X_1}(t) = \varphi_{X_2}(t)$ para cada número real t, entonces la función de distribución de X_2 debe ser F_{X_1}, ya que, como la siguiente proposición indica, la función de distribución se obtiene a partir de la función característica, por lo que, si ambas funciones características coinciden, también lo harán las funciones de distribución.

Veamos el resultado que, como ya hemos comentado, no vamos a demostrar.

Proposición 8.6 *Sea X una variable aleatoria y sean F y φ sus funciones de distribución y característica, respectivamente. Entonces, si x_1 y x_2 son puntos de continuidad de F,*

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{k\to\infty} \int_{-k}^{k} \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t)dt$$

Este resultado se denomina **fórmula** o **teorema de inversión**. A partir de él se obtiene un corolario importante llamado **teorema de unicidad**.

Corolario 8.6.1 *La función característica de una variable aleatoria determina de forma única su función de distribución.*

Demostración. Si en la expresión de la anterior proposición tomamos $x_2 = x$, donde x es un punto de continuidad de F, y hacemos tender x_1 a $-\infty$, puesto que

$$\lim_{x_1\to-\infty} F(x_1) = 0,$$

entonces,

$$F(x) = \frac{1}{2\pi} \lim_{x_1\to-\infty} \lim_{k\to\infty} \int_{-k}^{k} \frac{e^{-itx_1} - e^{-itx}}{it} \varphi(t)dt .$$

La expresión anterior demuestra que, conocida la función característica, queda determinada de forma única la función de distribución. ♣

Ejemplo 8.5. Se hacen 5 lanzamientos independientes de una moneda cuya probabilidad de cara es p. Cada tirada puede representarse por una variable aleatoria X_j, $j = 1, \ldots, 5$, que vale 1 cuando se obtiene una cara y, por tanto, con probabilidad p, y cero cuando se obtiene una cruz, con probabilidad $q = 1 - p$. La función característica de X_j se calcula con sencillez; en efecto,

$$\varphi_{X_j}(t) = e^{it\cdot 1} p[X_j = 1] + e^{it\cdot 0} p[X_j = 0] = p \cdot e^{it} + q .$$

Consideremos la variable aleatoria, X, *número de caras obtenidas en las cinco tiradas*; obviamente, $X = \sum_{j=1}^{5} X_j$, toma los valores $0, 1, 2, 3, 4$ y 5. Como se verá en el apartado 9.3, y el lector puede ya comprobar, su función de probabilidad vale

$$p[X = k] = \binom{5}{k} p^k q^{5-k} \qquad \text{si } k = 1, \ldots, 5 .$$

La función de probabilidad anterior depende del número de tiradas (cinco en este caso) y de la moneda (esto es, de p). Como veremos, asimismo, en el próximo capítulo, a este tipo de distribución se le denomina *distribución binomial de parámetros n y p*, donde n es el número de tiradas y p la probabilidad de obtener cara en una de ellas.

Obtengamos la función característica de X. Como $X_1, \ldots, X_5$ son independientes, aplicando el corolario 8.3.1,

$$\varphi_X(t) = \varphi_{\sum_{j=1}^{5} X_j}(t) = \left[\varphi_{X_j}(t)\right]^5 = (p \cdot e^{it} + q)^5 .$$

Obsérvese que el cálculo directo de la función característica de X resultaría más complicado, ya que tendríamos que hallar la siguiente esperanza:

$$\begin{aligned}\varphi_X(t) &= E(e^{itX}) = \sum_{k=1}^{n} e^{itk}\binom{5}{k}p^k q^{5-k} = \\ &= \sum_{k=1}^{n}\binom{5}{k}(pe^{it})^k q^{5-k} = (pe^{it}+q)^5 ,\end{aligned}$$

donde la última expresión se obtiene utilizando la fórmula de Newton para el desarrollo de la potencia de un binomio.

Un jugador lanza cinco veces esa moneda, y otro la lanza siete veces. Llamemos X al *número de caras obtenido por el primer jugador*, e Y al *número de caras obtenido por el segundo*. Sea Z el *número total de caras obtenido*, esto es, $Z = X + Y$. Hallemos la función característica de Z, suponiendo que los lanzamientos de ambos jugadores son independientes (el cálculo de la función característica de Y es análogo al de la de X). Así,

$$\begin{aligned}\varphi_Z(t) &= \varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t) = \\ &= (p\cdot e^{it}+q)^5\cdot(p\cdot e^{it}+q)^7 = (p\cdot e^{it}+q)^{12} .\end{aligned}$$

Si se compara esta función característica con la de X, se observa que son idénticas, salvo que en la de Z el número 5 se ha sustituido por un 12; por tanto, la función característica de Z es la de una distribución binomial con parámetros $n = 12$ y p. Aplicando el teorema de unicidad, la función de distribución de Z será la de una distribución binomial cuya correspondiente función de probabilidad es

$$p[Z=k] = \binom{5}{k}p^k q^{12-k} \qquad \text{si } k = 1,\ldots,12 .$$

En definitiva, si dos jugadores lanzan la misma moneda de forma independiente, la suma del número de caras obtenido por cada uno tiene el mismo comportamiento probabilístico (la misma función de probabilidad) que el número de caras obtenido al realizar un mismo jugador todas las tiradas. ◁

Veamos finalmente, a modo de síntesis, las relaciones existentes entre las funciones que describen o caracterizan cualquier distribución de probabilidades discreta o continua. La primera parte de cada uno de los dos enunciados incluidos en la siguiente proposición expresa cómo obtener la función característica a partir de la función de probabilidad o de la función de densidad, respectivamente. La segunda, que no demostraremos, se deduce del teorema de inversión (en el caso discreto es una aplicación inmediata, no así en el caso continuo, donde la demostración presenta cierta dificultad), e indica cómo hallar la función probabilidad o densidad mediante la función característica.

Proposición 8.7 *Sea X una variable aleatoria y sean f y φ sus funciones de probabilidad, o de densidad, y característica, respectivamente. Entonces,*

1) Si X es discreta:

- $\varphi(t) = \sum_{j=1}^{\infty} e^{itx_j} f(x_j) = \sum_{j=1}^{\infty} e^{itx_j} p[X = x_j]$, *para todo* $t \in \mathbb{R}$.

- *Por otro lado,*

$$f(x_j) = p[X = x_j] = \frac{1}{2\pi} \lim_{k\to\infty} \int_{-k}^{k} \frac{e^{-it(x_{j-1}+h)} - e^{-it(x_j+h)}}{it} \varphi(t)dt \,,$$

donde h *cumple que* $x_{j-1} + h < x_j$ *y* $x_j + h < x_{j+1}$.

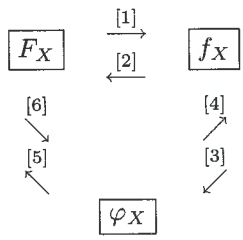
2) Si X *es continua:*

- $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx$, *para todo* $t \in \mathbb{R}$.
- *Además, si* $|\varphi|$ *es integrable en* $\mathbb{R}$,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \cdot \varphi_X(t)dt \,, \quad \textit{para todo } x \in \mathbb{R} \,.$$

En definitiva, este resultado, junto con la fórmula de inversión, describe las relaciones existentes entre función de distribución, función de probabilidad o de densidad y función característica. No sólo pueden obtenerse unas a partir de otras, sino que cualquiera de ellas caracteriza la distribución de probabilidades de una variable aleatoria discreta o continua.

El siguiente esquema, donde F_X, f_X y φ_X denotan las funciones de distribución, de probabilidad o de densidad y característica, respectivamente, de una variable aleatoria X, ilustra este comentario.



La relación [1] viene expresada, como ya sabemos, en la forma:

$$f_X(x_i) = p[X = x_i] = F_X(x_i) - F_X(x_{i-1}) \,,$$

si X es discreta, y

$$f_X(x) = \frac{dF_X(x)}{dx} \,,$$

si X es continua.

Su inversa, la relación [2], viene reflejada en las igualdades:

$$F_X(x) = \sum_{x_i \le x} f_X(x_i) = \sum_{x_i \le x} p[X = x_i]$$

y

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \,,$$

según que X sea discreta o continua.

Las relaciones [3] y [4] están recogidas en la proposición anterior. La relación [5] no es sino el corolario 8.6.1 del teorema de inversión y, por último, la [6] se deduce de la composición de las relaciones [1] y [3]. Cada uno de los tres tipos de funciones presenta ciertas ventajas y ciertos inconvenientes en su utilización.

Así, la función de distribución es la base conceptual de las distribuciones, teniendo un claro significado estadístico (la probabilidad que se va acumulando en la recta real); en cambio, su representación gráfica sólo proporciona una vaga idea de la distribución de la probabilidad, siendo, además, una función poco operativa.

Por su parte, la función de probabilidad (probabilidad de un valor de la variable) o la función de densidad (densidad de probabilidad en un intervalo *infinitesimal* de la recta real), además de tener un evidente significado estadístico, son muy operativas y sus representaciones gráficas dejan ver con claridad la distribución de la probabilidad.

Por otro lado, la función característica no tiene ningún significado conceptual y su *irrealizable* representación gráfica (¡en tres dimensiones!) no revelaría nada acerca de la distribución de la probabilidad. Sin embargo, sus propiedades matemáticas la convierten en un eficaz instrumento para abordar ciertos problemas (distribuciones de combinaciones lineales de variables aleatorias, cálculos de momentos, etc.) que hemos concretado en las proposiciones anteriores.

8.3 Función característica de una variable aleatoria n-dimensional. Propiedades

Desarrollemos esta sección considerando el caso general, n variables, ilustrando los resultados con ejemplos referidos a dos variables.

Definición 8.4 *Sea (Ω, α, p) un espacio probabilístico y sea $(X_1, \ldots, X_n)$ una variable aleatoria n-dimensional sobre Ω. Llamamos función característica de dicha variable n-dimensional a la aplicación*

$$\varphi_{(X_1,\ldots,X_n)} : \mathbb{R}^n \longrightarrow C ,$$

donde, para todo $(t_1, \ldots, t_n) \in \mathbb{R}^n$,

$$\varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = E[e^{i(t_1X_1+\cdots+t_nX_n)}] .$$

En el apartado 7.5 indicamos que ciertas expresiones referentes a variables aleatorias n-dimensionales se simplificaban si las escribíamos como matrices columna. Designando por $\mathbf{X}$ la matriz columna formada por las n variables, y por $\mathbf{t}$ la matriz

columna formada por n números reales cualesquiera, podemos expresar la definición anterior en forma matricial[1] como

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\left(e^{i\mathbf{t}'\mathbf{X}}\right) .$$

Denotaremos, asimismo, por $\mathbf{x}$ a la matriz columna formada por una realización de la variable n-dimensional $\mathbf{X}$, es decir, $\mathbf{x} = (x_1, \ldots, x_n)'$. Entonces, teniendo en cuenta que

$$e^{i\mathbf{t}'\mathbf{X}} = \cos(\mathbf{t}'\mathbf{X}) + i\operatorname{sen}(\mathbf{t}'\mathbf{X}) ,$$

y de acuerdo con la definición 8.2, la función característica de una variable n-dimensional, $\mathbf{X}$, valdrá:

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E[\cos(\mathbf{t}'\mathbf{X})] + i \cdot E[\operatorname{sen}(\mathbf{t}'\mathbf{X})] .$$

Por tanto, si $\mathbf{X}$ es una variable aleatoria n-dimensional discreta, su función característica será

$$\begin{aligned}
\varphi_{\mathbf{X}}(\mathbf{t}) &= \sum_{x_1} \cdots \sum_{x_n} \cos(\mathbf{t}'\mathbf{x}) \cdot p[X_1 = x_1, \ldots, X_n = x_n] + \\
&\quad + i \cdot \sum_{x_1} \cdots \sum_{x_n} \operatorname{sen}(\mathbf{t}'\mathbf{x}) \cdot p[X_1 = x_1, \ldots, X_n = x_n] = \\
&= \sum_{x_1} \cdots \sum_{x_n} [\cos(\mathbf{t}'\mathbf{x}) + i \cdot \operatorname{sen}(\mathbf{t}'\mathbf{x})] \cdot p[X_1 = x_1, \ldots, X_n = x_n] = \\
&= \sum_{x_1} \cdots \sum_{x_n} e^{i\mathbf{t}'\mathbf{x}} p[X_1 = x_1, \ldots, X_n = x_n] ,
\end{aligned}$$

o, lo que es lo mismo,

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \sum_{x_1} \cdots \sum_{x_n} e^{i(t_1 x_1 + \cdots + t_n x_n)} \cdot p[X_1 = x_1, \ldots, X_n = x_n] .$$

Si $\mathbf{X}$ es continua,

$$\begin{aligned}
\varphi_{\mathbf{X}}(\mathbf{t}) &= \int_{x_1} \cdots \int_{x_n} \cos(\mathbf{t}'\mathbf{x}) \cdot f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n + \\
&\quad + i \int_{x_1} \cdots \int_{x_n} \operatorname{sen}(\mathbf{t}'\mathbf{x}) \cdot f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n =
\end{aligned}$$

1. Obsérvese que el producto matricial $\mathbf{t}'\mathbf{X}$ multiplica la matriz fila $\mathbf{t}' = (t_1, \ldots, t_n)$ por la matriz columna aleatoria

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} .$$

El resultado es la matriz 1×1

$$\mathbf{t}'\mathbf{X} = t_1 X_1 + \cdots + t_n X_n ,$$

que se identifica de manera natural con la variable aleatoria unidimensional $t_1 X_1 + \cdots + t_n X_n$.

$$= \int_{x_1} \cdots \int_{x_n} [\cos(\mathbf{t}'\mathbf{x}) + i \cdot \operatorname{sen}(\mathbf{t}'\mathbf{x})] \cdot f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n =$$

$$= \int_{x_1} \cdots \int_{x_n} e^{i\mathbf{t}'\mathbf{x}} \cdot f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n ,$$

o, lo que es lo mismo,

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_{x_1} \cdots \int_{x_n} e^{i(t_1 x_1 + \cdots + t_n x_n)} \cdot f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n .$$

Adviértase que el cálculo de las funciones características de las variables n-dimensionales presenta los mismos problemas que en el caso unidimensional, agravadas aquí por la mayor dimensión.

Ejemplo 8.6. Consideremos una variable aleatoria (X, Y) cuya función de probabilidad viene recogida en la siguiente tabla:

$X\backslash Y$	-1	1	
0	$3/6$	$1/6$	$4/6$
1	$2/6$	0	$2/6$
	$5/6$	$1/6$	1

Calculemos su función característica

$$\varphi_{(X,Y)}(t_1, t_2) = E[e^{i(t_1 X + t_2 Y)}] = e^{i[t_1 0 + t_2(-1)]} \cdot \frac{3}{6} + e^{i[t_1 1 + t_2(-1)]} \cdot \frac{2}{6} + e^{i(t_1 0 + t_2 1)} \cdot \frac{1}{6} .$$

De aquí,

$$\varphi_{(X,Y)}(t_1, t_2) = \frac{1}{6}\left[3e^{-it_2} + 2e^{i(t_1 - t_2)} + e^{it_2}\right] . \triangleleft$$

Ejemplo 8.7. Sea (X, Y) una variable aleatoria bidimensional, con función de densidad conjunta

$$f(x, y) = \begin{cases} 1 & 0 < x < 1\ ,\ 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

Hallemos su función característica, suponiendo que las integrales dobles en el campo complejo pueden resolverse iterativamente como en el campo real (el lector interesado puede realizar el cálculo riguroso de la función característica, calculando las esperanzas de la parte real y de la parte imaginaria). Así, si t_1 y t_2 son distintos de cero,

$$\varphi_{(X,Y)}(t_1, t_2) = E\left[e^{i(t_1 X + t_2 Y)}\right] = \int_{x=0}^{1} \int_{y=0}^{1} e^{i(t_1 x + t_2 y)} dy dx =$$

$$= \int_{x=0}^{1} e^{it_1 x} dx \cdot \int_{y=0}^{1} e^{it_2 y} dy =$$

$$= \frac{e^{it_1} - 1}{it_1} \cdot \frac{e^{it_2} - 1}{it_2} ,$$

sin más que tener en cuenta la resolución de estas integrales que se realizó en el ejemplo 8.2.

De la misma forma, obtendríamos la función característica cuando t_1, t_2 o ambos son nulos. Así,

$$\varphi_{(X,Y)}(t_1,t_2)=\begin{cases}\dfrac{e^{it_1}-1}{it_1}\cdot\dfrac{e^{it_2}-1}{it_2} & t_1\neq 0, t_2\neq 0\\ \dfrac{e^{it_1}-1}{it_1} & t_1\neq 0, t_2=0\\ \dfrac{e^{it_2}-1}{it_2} & t_1=0, t_2\neq 0\\ 1 & t_1=0, t_2=0\end{cases}$$

resulta ser la función característica de (X,Y). ◁

Al igual que para las variables unidimensionales, también en este caso podemos hablar de la **función generatriz de momentos**, que se define como *una aplicación*

$$\phi_{(X_1,\ldots,X_n)}:\mathbb{R}^n\longrightarrow\mathbb{R}\,,$$

tal que, para todo $(t_1,\ldots,t_n)\in\mathbb{R}^n$,

$$\phi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)=E\left(e^{t_1X_1+\cdots+t_nX_n}\right)\,.$$

Esta función presenta el inconveniente de no tener asegurada su existencia, mientras que la función característica existe siempre, como el lector puede comprobar siguiendo un razonamiento análogo al que utilizamos en la proposición 8.1.

Las siguientes proposiciones demuestran propiedades de interés que verifican las funciones características de variables aleatorias n-dimensionales. La primera permite hallar, a partir de la función característica de una variable n-dimensional, la función característica de cualquiera de sus componentes, es decir, las funciones características marginales a partir de la función característica conjunta. Este resultado puede ampliarse, como veremos, para obtener la función característica de cualquier subgrupo de variables. La segunda proposición y su corolario relacionan la función característica de una combinación lineal de variables con la función característica conjunta de todas ellas. Estos resultados confirman de nuevo que, aunque la función característica tiene una gran complejidad formal, resulta sencillo operar con ella para calcular funciones características de ciertas variables.

Proposición 8.8 *Sea* $\varphi_{(X_1,\ldots,X_n)}$ *la función característica de la variable* n*-dimensional* $(X_1,\ldots,X_n)$. *Entonces,*

$$\varphi_{X_j}(t_j)=\varphi_{(X_1,\ldots,X_n)}(0,\ldots,t_j,\ldots,0)\,,$$

para $j=1,\ldots,n$.

Demostración. Resulta inmediata, sin más que sustituir el valor $(0,\ldots,t_j,\ldots,0)$ en la expresión general de la función característica n-dimensional. En efecto,

$$\varphi_{(X_1,\ldots,X_n)}(0,\ldots,t_j,\ldots,0)=E\left[e^{i(0\cdot X_1+\cdots+t_jX_j+\cdots+0\cdot X_n)}\right]=E\left(e^{it_jX_j}\right)=\varphi_{X_j}(t_j).\ \clubsuit$$

Igualmente, podríamos hallar la función característica de cualquier subconjunto de variables de $(X_1, \ldots, X_n)$. La regla general es que la función característica de un grupo de variables extraídas de la variable n-dimensional se obtiene a partir de la función característica de ésta, sustituyendo por ceros los valores t_i correspondientes a las variables que no intervienen. Por ejemplo, si disponemos de la función característica de la variable aleatoria (X_1, X_2, X_3, X_4), la de la variable bidimensional (X_1, X_3) se halla como

$$\varphi_{(X_1,X_3)}(t_1, t_3) = \varphi_{(X_1,X_2,X_3,X_4)}(t_1, 0, t_3, 0).$$

Históricamente, el uso de la función característica se generalizó dada su facilidad para estudiar la distribución de combinaciones lineales de variables aleatorias y, en concreto, de sus sumas. Veamos, en este sentido, el siguiente resultado.

Proposición 8.9 *Sea una variable aleatoria n-dimensional, $(X_1, \ldots, X_n)$, cuya función característica es $\varphi_{(X_1,\ldots,X_n)}$. Para cualesquiera números reales $a_1, \ldots, a_n$ se cumple que*

$$\varphi_{\sum_{j=1}^n a_j X_j}(t) = \varphi_{(X_1,\ldots,X_n)}(a_1 t, \ldots, a_n t) .$$

Demostración. Es inmediata, sin más que sustituir:

$$\begin{aligned} \varphi_{\sum_{j=1}^n a_j X_j}(t) &= E\left(e^{it\sum_{j=1}^n a_j X_j}\right) = \\ &= E\left[e^{i\sum_{j=1}^n (a_j t) X_j}\right] = \varphi_{(X_1,\ldots,X_n)}(a_1 t, \ldots, a_n t) . \clubsuit \end{aligned}$$

Así, obtendríamos la función característica de la suma de variables aleatorias a partir de la función característica conjunta, según se recoge en el siguiente corolario.

Corolario 8.9.1 *La función característica de la suma de n variables aleatorias se obtiene a partir de su función característica conjunta en la forma*

$$\varphi_{\sum_{j=1}^n X_j}(t) = \varphi_{(X_1,\ldots,X_n)}(t, \ldots, t) .$$

Basta con sustituir $a_1 = \cdots = a_n = 1$ en la proposición anterior.

Obsérvese la relación del corolario anterior con la proposición 8.3. En ésta obteníamos la función característica de la suma de variables independientes mediante las funciones características de cada una de las variables. El corolario anterior proporciona la función característica de la suma de n variables aleatorias cualesquiera (sean o no independientes), utilizando la función característica conjunta de todas ellas. De hecho, existe una sencilla relación entre la función característica conjunta de n variables aleatorias independientes y sus funciones características marginales, como muestra la siguiente proposición.

Proposición 8.10 *Las variables aleatorias $X_1, \ldots, X_n$ son independientes si y solamente si*

$$\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n) = \prod_{j=1}^{n} \varphi_{X_j}(t_j) \ ,$$

siendo $\varphi_{(X_1,\ldots,X_n)}$ la función característica de la variable aleatoria n-dimensional $(X_1,\ldots,X_n)$, y φ_{X_j} la de la variable X_j, con $j = 1,\ldots,n$.

Demostración. No daremos una demostración formal, sino sólo una ilustración, de este resultado. Veamos la condición necesaria. Si las variables $X_1,\ldots,X_n$ son independientes, lo serán también (este hecho no lo probaremos) las n variables aleatorias complejas $e^{it_1X_1},\ldots,e^{it_nX_n}$. Si aceptamos, asimismo, que para variables complejas independientes la esperanza de su producto es el producto de sus esperanzas,

$$\begin{aligned}\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n) &= E\left[e^{i(t_1X_1+\cdots+t_nX_n)}\right] = E(e^{it_1X_1}\cdots e^{it_nX_n}) = \\ &= E(e^{it_1X_1})\cdots E(e^{it_nX_n}) = \varphi_{X_1}(t_1)\cdots\varphi_{X_n}(t_n) \ .\end{aligned}$$

La demostración de la condición suficiente, que no realizaremos, está basada en las fórmulas de inversión que enunciaremos más adelante. ♣

En el capítulo 5 hemos caracterizado la independencia de variables aleatorias a partir de la función de distribución, de la función de probabilidad o de densidad (en ambos casos, la caracterización indicaba que la función conjunta es igual al producto de las marginales). Como muestra el resultado anterior, esta caracterización de independencia se repite con la función característica.

No debemos confundir el resultado que proporciona la proposición 8.10 anterior con el establecido en la proposición 8.3, que señalaba que la función característica de la suma de variables independientes es el producto de las funciones características de los sumandos. Este resultado no caracteriza la independencia de variables aleatorias, mientras que la descomposición que enunciamos en la proposición 8.10 sí lo hace. Ilustremos esta idea para dos variables.

Sea (X,Y) una variable aleatoria bidimensional. La proposición 8.3 indica que, si X e Y son independientes,

$$\varphi_{X+Y}(t) = \varphi_X(t)\cdot\varphi_Y(t).$$

Por otro lado, en la proposición 8.10 se demuestra que, si X e Y son independientes, su función característica conjunta se descompone como producto de las marginales, esto es,

$$\varphi_{(X,Y)}(t_1,t_2) = \varphi_X(t_1)\cdot\varphi_Y(t_2).$$

Obsérvese que esta última es una descomposición de una función de dos variables, (t_1,t_2), mientras que la primera lo es de una función de una única variable, t. Además, la última descomposición es más *fuerte*, esto es, implica la primera. Veamos este resultado.

Para dos variables aleatorias se cumple, tanto si son independientes como si no lo son, que

$$\varphi_{X+Y}(t) = \varphi_{(X,Y)}(t,t) \ ;$$

además, si las variables son independientes,

$$\varphi_{X+Y}(t) = \varphi_{(X,Y)}(t,t) = \varphi_X(t) \cdot \varphi_Y(t).$$

No es de extrañar, por tanto, que la segunda descomposición caracterice la independencia de variables aleatorias y no así la primera. En concreto, la proposición 8.10 indica que, si dos variables X e Y cumplen la condición

$$\varphi_{(X,Y)}(t_1,t_2) = \varphi_X(t_1) \cdot \varphi_Y(t_2),$$

para todo punto del plano (t_1,t_2), entonces X e Y son independientes. Pero, si dos variables aleatorias verifican que

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t),$$

no son necesariamente independientes, según ilustra el ejercicio 8.14. Obsérvese, con todo, que la situación propuesta en dicho ejercicio es muy particular.

Como en el caso unidimensional, pueden calcularse los momentos respecto del origen a partir de la función característica. El resultado, que es una extensión del que demostramos en la proposición 8.4, no lo probaremos, aunque las pautas de la demostración son las mismas que las de la proposición citada.

Proposición 8.11 *Sea $\varphi_{(X_1,\ldots,X_n)}$ la función característica de una variable aleatoria n-dimensional, $(X_1,\ldots,X_n)$. Si existe el momento no central de orden j_1 para X_1, ..., de orden j_n para X_n, la función característica es derivable j_1 veces respecto de t_1, ... , j_n veces respecto de t_n, en el punto $(t_1,\ldots,t_n) = (0,\ldots,0)$, y se cumple que*

$$E(X_1^{j_1} \cdots X_n^{j_n}) = \frac{1}{i^{j_1+\cdots+j_n}} \cdot \frac{\partial^{j_1+\cdots+j_n}}{\partial t_1^{j_1} \ldots \partial t_n^{j_n}} \varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)\Big|_{\mathbf{t=0}},$$

donde $\mathbf{0}$ *es la matriz columna formada por n ceros.*

A modo de ilustración, demostraremos este resultado únicamente para hallar $E(X_1 \cdot X_2)$ en una variable aleatoria bidimensional, a partir de su función característica,

$$\varphi_{(X_1,X_2)}(t_1,t_2) = E\left[e^{i(t_1X_1+t_2X_2)}\right].$$

Entonces,

$$\frac{\partial^2}{\partial t_1 \partial t_2} \varphi_{(X_1,X_2)}(t_1,t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} E\left[e^{i(t_1X_1+t_2X_2)}\right] = E\left[i^2 X_1 X_2 e^{i(t_1X_1+t_2X_2)}\right],$$

resultado obtenido derivando dentro de la serie que define la esperanza, si X es discreta, o derivando bajo el signo integral, si X es continua.

Para $(t_1,t_2) = (0,0)$,

$$\frac{\partial^2}{\partial t_1 \partial t_2} \varphi_{(X_1,X_2)}(t_1,t_2)\Big|_{(t_1,t_2)=(0,0)} = i^2 \cdot E(X_1 \cdot X_2),$$

de donde,

$$E(X_1 \cdot X_2) = \frac{1}{i^2} \cdot \frac{\partial^2}{\partial t_1 \partial t_2} \varphi_{(X_1,X_2)}(t_1,t_2)\Big|_{(t_1,t_2)=(0,0)} =$$
$$= -\frac{\partial^2}{\partial t_1 \partial t_2} \varphi_{(X_1,X_2)}(t_1,t_2)\Big|_{(t_1,t_2)=(0,0)}$$

Como los momentos no centrales de cada una de las variables pueden considerarse como momentos no centrales de la variable n-dimensional, pueden hallarse a partir de la expresión de la proposición 8.11, ya que, por ejemplo,

$$E(X_1) = E(X_1^1 X_2^0 \cdots X_n^0) = \frac{1}{i} \cdot \frac{\partial}{\partial t_1} \varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)\Big|_{\mathbf{t=0}}$$

y

$$E(X_2^5) = E(X_1^0 X_2^5 \cdots X_n^0) = \frac{1}{i^5} \cdot \frac{\partial^5}{\partial t_2^5} \varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)\Big|_{\mathbf{t=0}}$$

No obstante, resulta más sencillo su cálculo mediante las funciones características marginales.

La función característica de una variable n-dimensional caracteriza su distribución de probabilidades, como lo hacen la función de probabilidad o de densidad y la función de distribución n-dimensional. Al igual que en el caso unidimensional, estas tres funciones se relacionan, pudiéndose hallar una de ellas a partir de cualquiera de las otras dos. Dejamos al lector que explicite estas relaciones y sólo especificaremos el teorema de inversión, que nos permite calcular la función de distribución utilizando la función característica.

Proposición 8.12 *Sea $(X_1,\ldots,X_n)$ una variable aleatoria n-dimensional y sea φ su función característica. Entonces, para n pares de números a_i y b_i $(a_i < b_i)$,*

$$p[a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n] =$$
$$= \tfrac{1}{(2\pi)^n} \lim_{k\to\infty} \int_{-k}^{k} \cdots \int_{-k}^{k} \varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n) \prod_{j=1}^{n} \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j}\, dt_1 \ldots dt_n.$$

Como recordará el lector, si tomamos límites en la probabilidad anterior cuando a_1, ..., a_n tienden a $-\infty$, el resultado es la función de distribución en el punto $(b_1,\ldots,b_n)$. En definitiva, la proposición anterior, junto con este comentario, garantizan que la función característica determina la función de distribución. Por tanto, la función característica caracteriza la distribución de probabilidades de una variable n-dimensional.

Ejemplo 8.8. Consideremos, de nuevo, la variable aleatoria (X,Y) del ejemplo 8.6, cuya función de probabilidad es:

$X\backslash Y$	-1	1	
0	$3/6$	$1/6$	$4/6$
1	$2/6$	0	$2/6$
	$5/6$	$1/6$	1

Su función característica, ya calculada, es

$$\varphi_{(X,Y)}(t_1,t_2)=\frac{1}{6}\left[3e^{-it_2}+2e^{i(t_1-t_2)}+e^{it_2}\right] .$$

Si anulamos, o bien t_1, o bien t_2, obtendremos las funciones características marginales. Así, haciendo $t_2=0$,

$$\varphi_X(t_1)=\varphi_{(X,Y)}(t_1,0)=\frac{3+2e^{it_1}+1}{6}=\frac{2}{3}+\frac{1}{3}e^{it_1} .$$

Obsérvese que coincide con la función característica hallada en el ejemplo 8.1. Análogamente, la función característica de Y es

$$\varphi_Y(t_2)=\varphi_{(X,Y)}(0,t_2)=\frac{1}{6}\cdot(5e^{-it_2}+e^{it_2}) ,$$

función característica que podríamos haber calculado, también, sin más que considerar la distribución marginal de Y.

Evidentemente,

$$\varphi_{(X,Y)}(t_1,t_2)\neq\varphi_X(t_1)\cdot\varphi_Y(t_2) ,$$

es decir, X e Y son variables dependientes (resultado que, asimismo, puede comprobarse en la tabla de doble entrada que recoge las funciones de probabilidad conjunta y marginales).

De forma sencilla, es posible obtener, por ejemplo, la función característica de la variable $Z=3X-2Y$. Así,

$$\varphi_Z(t)=\varphi_{3X-2Y}(t)=\varphi_{(X,Y)}(3t,-2t)=\frac{1}{6}\cdot(3e^{2it}+2e^{5it}+e^{-2it}) .$$

El lector puede comprobar el resultado a partir de la distribución de Z, es decir,

$$Z:\left|\begin{array}{lll} -2 & p[Z=-2]=p[X=0,Y=1]=\dfrac{1}{6} \\ 2 & p[Z=2]=p[X=0,Y=-1]=\dfrac{3}{6} \\ 5 & p[Z=5]=p[X=1,Y=-1]=\dfrac{2}{6} \end{array}\right.$$

Por último, calculemos la covarianza entre X e Y, derivando la función característica conjunta (el lector advertirá que los cálculos, en este caso, son mucho más complicados que si se hiciesen directamente). Así,

$$\begin{aligned} E(X\cdot Y) &= \frac{1}{i^2}\cdot\left.\frac{\partial^2\varphi_{(X,Y)}(t_1,t_2)}{\partial t_1\partial t_2}\right|_{(t_1,t_2)=(0,0)}= \\ &= \frac{1}{i^2}\cdot\left(-\frac{1}{6}\right)i^2 2e^{i(t_1-t_2)}\Big|_{(t_1,t_2)=(0,0)}=-\frac{1}{3} , \end{aligned}$$

$$E(X)=\frac{1}{i}\cdot\left.\frac{\partial\varphi_{(X,Y)}(t_1,t_2)}{\partial t_1}\right|_{(t_1,t_2)=(0,0)}=\frac{1}{i}\cdot\frac{1}{6}i2e^{i(t_1-t_2)}\Big|_{(t_1,t_2)=(0,0)}=\frac{1}{3} \qquad \text{y}$$

$$\begin{aligned} E(Y) &= \frac{1}{i}\cdot\left.\frac{\partial\varphi_{(X,Y)}(t_1,t_2)}{\partial t_2}\right|_{(t_1,t_2)=(0,0)}= \\ &= \frac{1}{i}\cdot\frac{1}{6}\left[-3ie^{-it_2}-2ie^{i(t_1-t_2)}+ie^{it_2}\right]\Big|_{(t_1,t_2)=(0,0)}=-\frac{2}{3} . \end{aligned}$$

De aquí,

$$Cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y) = -\frac{1}{9} \,. \triangleleft$$

Veamos, ahora, un ejemplo para una variable bidimensional continua.

Ejemplo 8.9. Dada la variable bidimensional (X,Y) con función de densidad conjunta

$$f(x,y) = \begin{cases} 1 & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

en el ejemplo 8.7 calculamos su función característica:

$$\varphi_{(X,Y)}(t_1,t_2) = \begin{cases} \dfrac{e^{it_1}-1}{it_1} \cdot \dfrac{e^{it_2}-1}{it_2} & t_1 \neq 0, t_2 \neq 0 \\ \dfrac{e^{it_1}-1}{it_1} & t_1 \neq 0, t_2 = 0 \\ \dfrac{e^{it_2}-1}{it_2} & t_1 = 0, t_2 \neq 0 \\ 1 & t_1 = 0, t_2 = 0 \end{cases}$$

Anulando t_2 y t_1, obtenemos las funciones características de X y de Y, respectivamente. Así,

$$\varphi_X(t_1) = \varphi_{(X,Y)}(t_1,0) = \begin{cases} \frac{e^{it_1}-1}{it_1} & t_1 \neq 0 \\ 1 & t_1 = 0 \end{cases}$$

y

$$\varphi_Y(t_2) = \varphi_{(X,Y)}(0,t_2) = \begin{cases} \frac{e^{it_2}-1}{it_2} & t_2 \neq 0 \\ 1 & t_2 = 0 \end{cases}$$

Obviamente, dichas funciones características son idénticas; por tanto, X e Y presentan la misma distribución de probabilidad. En el ejemplo 8.2 ya habíamos calculado esta función característica a partir de la función de densidad de X.

Ahora, podemos comprobar que

$$\varphi_{(X,Y)}(t_1,t_2) = \varphi_X(t_1) \cdot \varphi_Y(t_2) \,,$$

es decir, X e Y son variables independientes.

Dado que X e Y son independientes, puede hallarse la función característica de su suma, bien utilizando su función característica conjunta, bien utilizando las marginales. Así,

$$\varphi_{X+Y}(t) = \varphi_{(X,Y)}(t,t) = \begin{cases} \frac{e^{it}-1}{it} \cdot \frac{e^{it}-1}{it} = \left(\frac{e^{it}-1}{it}\right)^2 & t \neq 0 \\ 1 & t = 0 \end{cases}$$

o bien,

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_X(t) = [\varphi_X(t)]^2 = \begin{cases} \left(\frac{e^{it}-1}{it}\right)^2 & t \neq 0 \\ 1 & t = 0 \end{cases}$$

En este ejemplo, hemos considerado conjuntamente dos variables aleatorias independientes e igualmente distribuidas. Posteriores desarrollos de la Estadística evidenciarán la importancia de este tipo de distribuciones conjuntas. $\triangleleft$

8.4 Ejercicios propuestos

8.1. Sea φ_X la función característica de una variable aleatoria unidimensional. Demuéstrese que:

a) $\varphi_X(0) = 1$.

b) $|\varphi_X(t)| \leq 1$.

c) $\varphi_X(-t)$ es el conjugado de $\varphi_X(t)$.

8.2. Sea X una variable aleatoria que toma siempre el valor 2. Calcúlese su función característica.

8.3. De una urna compuesta por 3 bolas blancas y 2 negras, una persona extrae tres bolas con devolución. Se pide:

a) Función característica de X, número de bolas negras extraídas hasta obtener una bola blanca.

b) Esperanza y varianza de X, a partir del resultado anterior.

8.4. Sea X una variable aleatoria continua cuya función de densidad es

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

Hállese la función característica de X.

8.5. Para contratar a sus mecanógrafas, una empresa realiza entre las aspirantes un examen consistente en escribir a máquina por espacio de una hora. El sueldo que asigna a las elegidas está en función del número de folios que han completado en dicho tiempo. Según este criterio establece la siguiente clasificación:

- Las secretarias de clase A, con más de ocho folios por hora, recibirán un sueldo de 1000 pesetas por hora.
- Las secretarias de clase B, entre seis y ocho folios por hora, con un sueldo de 700 pesetas por hora.
- Las secretarias de clase C, de cuatro a seis folios por hora, que percibirán un sueldo de 500 pesetas por hora.

Si admitimos que el número de folios que escribe una secretaria es una variable aleatoria, X, con función de densidad:

$$f(x) = \begin{cases} \frac{1}{6} & 4 < x < 10 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese la función característica de la variable aleatoria, Y, salario por hora de una mecanógrafa en esta empresa.

b) Obténgase el salario medio por hora de una secretaria a partir del resultado del apartado anterior.

8.6. Ana está muy disgustada por el servicio que se le ha prestado en la Seguridad Social ya que ha tenido que esperar tres meses para ser sometida a una cierta operación.
Si el tiempo, en meses, que un paciente en las condiciones de Ana tiene que aguardar hasta ser intervenido quirúrgicamente es una variable aleatoria cuya función característica es

$$\varphi_X(t) = (1 - 4it)^{-1} ,$$

¿tiene Ana motivos para estar enfadada?

8.7. Sea φ_X la función característica de una variable unidimensional. Demuéstrese que

$$Var(X) = \frac{1}{i^2} \cdot \left. \frac{\partial^2 \ln \varphi_X(t)}{\partial t^2} \right|_{t=0} ,$$

teniendo en cuenta que $\varphi_X(0) = 1$.

8.8. Sea $\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)$ la función característica de una variable aleatoria n-dimensional. Pruébese que:

a) $\varphi_{(X_1,\ldots,X_n)}(0,\ldots,0) = 1$.
b) $|\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)| \leq 1$.
c) $\varphi_{(X_1,\ldots,X_n)}(-t_1,\ldots,-t_n)$ es el conjugado de $\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)$.

8.9. Demuéstrese que la función característica de una variable aleatoria n-dimensional siempre existe.

8.10. De una urna compuesta por 3 bolas blancas y 2 negras, una persona extrae tres bolas con devolución. Se pide:

a) Función característica conjunta de X, número de bolas blancas extraídas, y de Y, número de bolas negras hasta obtener una blanca.
b) Funciones características marginales de X y de Y.
c) Calcúlese, a partir de las funciones características obtenidas en los apartados anteriores, el coeficiente de correlación entre X e Y.
d) De una urna idéntica a la anterior, otra persona extrae tres bolas con devolución. Obténgase la función característica de la variable número de bolas blancas extraídas entre las dos personas.

8.11. Sea (X, Y) una variable aleatoria bidimensional continua cuya función de densidad es

$$f(x,y) = \begin{cases} 3x & 0 \leq x \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

a) Hállese la función característica conjunta.
b) Calcúlense las funciones características marginales.
c) ¿Son X e Y independientes?
d) Obténgase la función característica de $Z = 2X - 3Y$.

8.12. A instancias de la Dirección General de Tráfico, los ayuntamientos de las ciudades con mayor parque automovilístico han elaborado informes sobre la seguridad vial en sus calles. A partir de dichos informes, se ha obtenido que la distribución conjunta de X, número de atropellos al mes en la vía con más densidad de tráfico de una cierta ciudad, y de Y, número de atropellos en la segunda vía más conflictiva, en cuanto a densidad, se modeliza según la siguiente función característica:

$$\varphi_{(X,Y)}(t_1,t_2)=\frac{e^{2(e^{it_1}-1)}}{e^{1-e^{it_2}}}\,.$$

a) Obténgase la función característica de cada una de las variables.
b) ¿Cuál de las dos calles tiene un mayor número medio mensual de atropellos?
c) Hállese una medida de la variabilidad del número de accidentes al mes en la vía con mayor densidad automovilística.
d) ¿Existe relación entre el número de accidentes por atropello en cada una de las dos calles?

8.13. Sea $\varphi_{(X,Y)}(t_1,t_2)$ la función característica de la variable (X,Y). Calcúlese, a partir de ella, la correspondiente a $X-Y$.

8.14. Sea X una variable aleatoria continua con función de densidad

$$f_X(x)=\frac{1}{\pi}\frac{1}{1+x^2},\quad -\infty<x<\infty\,.$$

Se demuestra que su función característica vale

$$\varphi_X(t)=e^{-|t|}\,.$$

Obténgase la función característica de $Y=2X$, y teniendo en cuenta que $X+Y=X+2X=3X$, compruébese que

$$\varphi_{X+Y}(t)=\varphi_X(t)\cdot\varphi_Y(t),$$

para todo t.
Pruébese, por otro lado, que X e Y no son independientes.

8.15. Sea $\varphi_{(X,Y)}$ la función característica de una variable bidimensional. Considerando que $\varphi_{(X,Y)}(0,0)=1$, demuéstrese que

a)

$$\alpha_{r,s}=E(X^r\cdot Y^s)=\frac{1}{i^{r+s}}\cdot\left.\frac{\partial^{r+s}\varphi_{(X,Y)}(t_1,t_2)}{\partial t_1^r\partial t_2^s}\right|_{(t_1,t_2)=(0,0)}\,.$$

b)

$$Cov(X,Y)=\frac{1}{i^2}\cdot\left.\frac{\partial^2\ln\varphi_{(X,Y)}(t_1,t_2)}{\partial t_1\partial t_2}\right|_{(t_1,t_2)=(0,0)}\,.$$

8.16. Sea $(X_1,\ldots,X_n)$ una variable aleatoria n-dimensional cuya función característica es el producto de las marginales, φ_{X_i}, para $i=1,\ldots,n$, esto es,

$$\varphi_{(X_1,\ldots,X_n)}(t_1,\ldots,t_n)=\varphi_{X_1}(t_1)\cdots\varphi_{X_n}(t_n)\,.$$

Pruébese que $X_1,\ldots,X_n$ son independientes.

Capítulo 9

MODELOS DE DISTRIBUCIONES DISCRETAS

Uno de los objetivos del Cálculo de Probabilidades es determinar distribuciones que puedan servir de modelos a los distintos fenómenos aleatorios que se presentan en las diversas ramas del saber humano. Muchas variables aleatorias asociadas con experimentos estadísticos tienen propiedades similares y pueden describirse esencialmente con la misma distribución de probabilidad. Esto ha llevado a la elaboración de una serie de **distribuciones tipo** en las cuales encuadraremos la mayoría de las variables aleatorias que se encuentran en la práctica. Por abuso del lenguaje, designaremos a las variables con la misma denominación que la de su distribución tipo. Se denominan distribuciones tipo porque la determinación de sus funciones de probabilidad o de densidad, según los casos, depende de los valores que asignemos a ciertas constantes desconocidas, constantes que se denominan **parámetros**.

Por ejemplo, si lanzamos una moneda y desconocemos su grosor, podríamos definir una variable aleatoria, X, que describiera los tres resultados posibles; X valdría la unidad, si se obtiene una cara; cero, si el resultado es una cruz, y 0.5 si la moneda cae de canto. Su función de probabilidad podría, entonces, escribirse de la siguiente forma:

$$X: \left| \begin{array}{ll} 1 & p[X=1]=p \\ 0.5 & p[X=0.5]=r \\ 0 & p[X=0]=q \end{array} \right.$$

Esta función de probabilidad depende, en principio, de tres parámetros, p, q y r. Realmente, sólo depende de dos de ellos porque, como los valores de la función de probabilidad tienen que sumar la unidad, se cumplirá que

$$p+q+r=1$$

y, por tanto,

$$q=1-p-r\,.$$

En definitiva, la función de probabilidad depende, únicamente, de los valores de p y r.

Asignando distintos valores a los parámetros, pueden obtenerse funciones de probabilidad que describan el comportamiento aleatorio de diferentes monedas. Por ejemplo, si $p = 0.2$ y $r = 0.6$, tendríamos $q = 0.2$; se trataría de una moneda *tubo* (con un gran grosor) no cargada, esto es, con las mismas probabilidades para los sucesos *obtener cara* y *obtener cruz*. Con $r = 0$ describiríamos el comportamiento aleatorio de las monedas usuales, para las que no se acepta que puedan caer de canto.

Cuando se intenta asociar a un fenómeno aleatorio un modelo que explique su comportamiento probabilístico, la práctica habitual consiste en asignarle una distribución tipo que sea razonable; a continuación, se buscan los valores de los parámetros que mejor se ajusten a la información disponible sobre el fenómeno.

En este capítulo y en los tres siguientes, se estudiarán algunas de estas distribuciones para variables aleatorias unidimensionales y n-dimensionales, tanto en el caso discreto como en el continuo. Para todas ellas se propone o se calcula su función de probabilidad o de densidad, según los casos. Posteriormente, se obtienen sus características más importantes (fundamentalmente, media, varianza y función característica) y se demuestran propiedades de interés. Muchos de estos cálculos y demostraciones son ejercicios que bien hubieran podido ser incluidos como tales en anteriores capítulos. En este sentido, sugerimos al lector que intente realizarlas sin la ayuda del libro, lo que le permitirá comprobar el grado de comprensión que de los capítulos anteriores ha alcanzado.

9.1 Distribución uniforme discreta

Si una variable aleatoria X toma N valores, $x_1, \ldots, x_N$, con la misma probabilidad, se dice que X sigue una **distribución uniforme discreta** de parámetro N, para $N = 1, 2, \ldots$; su función de probabilidad es, por tanto, $p[X = x_i] = k$, para todo $i = 1, \ldots, N$.

El valor de la constante k se calcula con facilidad, pues la suma de todos los valores de la función de probabilidad ha de ser igual a la unidad, esto es,

$$\sum_{i=1}^{N} p[X = x_i] = \sum_{i=1}^{N} k = N \cdot k = 1 \, ;$$

por tanto, $k = \frac{1}{N}$. La **función de probabilidad** de una distribución uniforme discreta queda, así, perfectamente determinada:

$$f_X(x_i) = p[X = x_i] = \frac{1}{N} \quad \text{para} \quad i = 1, \ldots, N \, .$$

Si suponemos ordenados los valores de la variable, esto es, $x_1 < x_2 < \cdots < x_N$, la representación gráfica de dicha función de probabilidad es la mostrada en la figura 9.1.

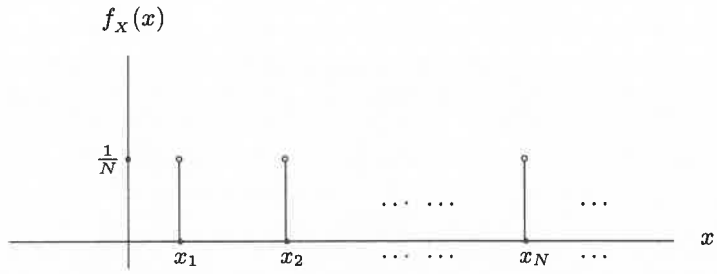


FIG. 9.1.

Un ejemplo sencillo de variable aleatoria con distribución uniforme discreta es la variable, X, que describe el resultado de la tirada de un dado perfecto. Para esta variable, $p[X = 1] = \cdots = p[X = 6] = \frac{1}{6}$.

Por acumulación de la función de probabilidad puede obtenerse la **función de distribución**, que resulta ser:

$$F_X(x) = \begin{cases} 0 & x < x_1 \\ \frac{1}{N} & x_1 \leq x < x_2 \\ \vdots & \vdots \\ \frac{k}{N} & x_k \leq x < x_{k+1} \\ \vdots & \vdots \\ \frac{N-1}{N} & x_{N-1} \leq x < x_N \\ 1 & x \geq x_N \end{cases}$$

Esta función es, por tanto, escalonada, siendo todos los *saltos* de la misma cuantía.

A continuación, calculemos las principales características de esta variable sin más que aplicar sus definiciones.

Esperanza:

$$E(X) = \sum_{i=1}^{N} x_i \cdot p[X = x_i] = \sum_{i=1}^{N} x_i \cdot \frac{1}{N} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \, .$$

Varianza:

$$E(X^2) = \sum_{i=1}^{N} x_i^2 \cdot p[X = x_i] = \sum_{i=1}^{N} x_i^2 \cdot \frac{1}{N} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2$$

y, por tanto,

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N} \cdot \sum_{i=1}^{N} x_i \right)^2 \, .$$

Función característica:

$$\varphi_X(t) = E(e^{itX}) = \sum_{j=1}^{N} e^{itx_j} \cdot p[X = x_j] =$$
$$= \sum_{j=1}^{N} e^{itx_j} \cdot \frac{1}{N} = \frac{1}{N} \cdot \sum_{j=1}^{N} e^{itx_j} .$$

Dos casos particulares de esta distribución, que describiremos seguidamente, son especialmente frecuentes en las aplicaciones.

Caso 1: Cuando $N = 1$, es decir, cuando la variable aleatoria toma un único valor (que llamaremos x_1) con probabilidad uno, la distribución uniforme recibe el nombre de degenerada, y una variable aleatoria que tenga esa distribución de probabilidad se denomina variable con distribución uniforme degenerada o simplemente distribución degenerada. Su función de probabilidad es

$$p[X = x_1] = 1$$

y la de distribución,

$$F_X(x) = \begin{cases} 0 & x < x_1 \\ 1 & x \geq x_1 \end{cases}$$

Sus momentos de orden uno y dos y su función característica son, respectivamente,

$$E(X) = x_1 \cdot 1 = x_1$$

y

$$E(X^2) = x_1^2 \cdot 1 = x_1^2 ,$$

y, por tanto,

$$Var(X) = x_1^2 - (x_1)^2 = 0 ;$$

finalmente,

$$\varphi_X(t) = E(e^{itX}) = e^{itx_1} \cdot 1 = e^{itx_1} .$$

Las variables degeneradas no son sino constantes, esto es, números reales. Sin embargo, en ciertas situaciones interesa considerarlas como variables aleatorias. El artificio consiste, pues, en asignarles una distribución degenerada.

Caso 2: Supongamos que $x_1 = 1, x_2 = 2, \ldots, x_N = N$; entonces, podemos escribir la función de probabilidad como

$$p[X = k] = \frac{1}{N}, \qquad \text{para } k = 1, \ldots, N .$$

La expresión de su función de distribución será, en consecuencia,

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{N} & 1 \leq x < 2 \\ \vdots & \vdots \\ \frac{k}{N} & k \leq x < k+1 \\ \vdots & \vdots \\ \frac{N-1}{N} & N-1 \leq x < N \\ 1 & x \geq N \end{cases}$$

Sus principales características se hallan sin más que concretar, para este caso, los resultados obtenidos para una variable uniforme discreta general[1].

Esperanza:

$$E(X) = \sum_{k=1}^{N} k \cdot p[X = k] = \sum_{k=1}^{N} k \cdot \frac{1}{N} = \frac{1}{N} \cdot \frac{N+1}{2} \cdot N = \frac{N+1}{2} .$$

Varianza:

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{N} k^2 \cdot \frac{1}{N} = \frac{1}{N} \cdot \sum_{k=1}^{N} k^2 = \frac{1}{N} \cdot \frac{N(N+1)(2N+1)}{6} = \\ &= \frac{(N+1)(2N+1)}{6} \end{aligned}$$

y, por tanto,

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 = \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 = \\ &= \frac{N^2-1}{12} . \end{aligned}$$

Función característica:

$$\begin{aligned} \varphi_x(t) &= E(e^{itX}) = \sum_{k=1}^{N} e^{itk} \cdot \frac{1}{N} = \frac{1}{N} \cdot \frac{e^{it} - e^{itN} \cdot e^{it}}{1 - e^{it}} = \\ &= \frac{e^{it}(1 - e^{itN})}{N(1 - e^{it})} . \end{aligned}$$

1. Recordemos que la suma de los n primeros términos de una progresión aritmética es

$$S = \frac{a_1 + a_n}{2} \cdot n ;$$

y que la suma de los n primeros términos de una progresión geométrica de razón r es

$$S = \frac{a_1 - a_n \cdot r}{1 - r} .$$

Por otra parte, la suma de los cuadrados de los n primeros números naturales es

$$1^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} .$$

9.2 Distribución de Bernoulli

Sea Ω el espacio muestral asociado a un experimento aleatorio. Con frecuencia, no interesa estudiar cuál ha sido el resultado concreto, ω, del experimento, sino sólo si ha ocurrido, o no, un cierto suceso, $A \in \Omega$, esto es, si $\omega \in A$. Decimos, entonces, que se trata de un **experimento de Bernoulli**.

En un experimento de Bernoulli, se denomina *éxito* al suceso en estudio, A, y *fracaso* a su contrario, $\overline{A}$. A dicho experimento le asociamos una variable aleatoria X definida como el *número de éxitos al realizar el experimento*, es decir,

$$X : \left| \begin{array}{ll} 0 & \text{si ocurre } \overline{A} \\ 1 & \text{si ocurre } A \end{array} \right.$$

Si llamamos $p(A) = p$ y $p(\overline{A}) = 1 - p = q$, la función de probabilidad de la variable X puede escribirse como

$$X : \left| \begin{array}{ll} 0 & p[X = 0] = 1 - p = q \\ 1 & p[X = 1] = p \end{array} \right.$$

Diremos que X tiene una **distribución de Bernoulli** de parámetro p, donde p es la probabilidad de éxito, que se encontrará entre 0 y 1 ($0 \leq p \leq 1$). La denotaremos brevemente como

$$X \hookrightarrow b(p) \, .$$

Su **función de probabilidad** puede expresarse de forma más compacta:

$$f_X(x) = p[X = x] = p^x \cdot (1 - p)^{1-x} \, , \text{ para } x = 0, 1 \, .$$

La representación gráfica de dicha función es la mostrada en la figura 9.2.

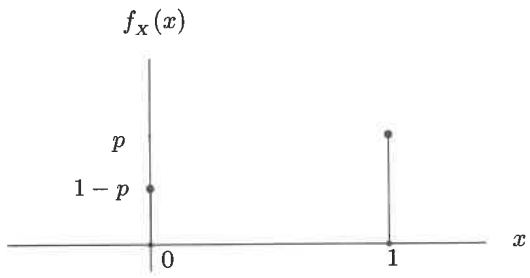


FIG. 9.2.

Obviamente, si p es igual a cero o a la unidad, la variable de Bernoulli es degenerada; si p vale cero, X vale cero con probabilidad igual a la unidad (no es posible obtener un éxito), y si p vale la unidad es seguro que se obtendrá un éxito (X vale la unidad con probabilidad uno).

Un sencillo ejemplo de variable de Bernoulli es el número de caras que se obtienen al lanzar una moneda al aire una vez. El parámetro, p, sería, en este caso, la probabilidad de obtener cara con dicha moneda.

Su **función de distribución** es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Los momentos y función característica de esta variable se calculan de modo sencillo. Su **esperanza** vale

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p \,,$$

esto es, la probabilidad de éxito; por lo que se refiere a la **varianza**,

$$E(X^2) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$

y, por tanto,

$$Var(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = p \cdot q \,.$$

Por su parte, la **función característica** es

$$\varphi_X(t) = E(e^{itX}) = e^{it0} \cdot (1-p) + e^{it1} \cdot p = q + p \cdot e^{it} \,.$$

Ejemplo 9.1. Un posible experimento de Bernoulli sería el lanzamiento de un tiro libre por parte de un jugador de baloncesto, ya que los dos posibles resultados son *conseguir canasta* (éxito) y *no conseguir canasta* (fracaso). Sabemos, además, que, según se deduce de los partidos de liga que ha disputado hasta el momento, dicho jugador tiene aproximadamente un 80 % de aciertos en tiros libres. Este conocimiento nos permite asociar al experimento la variable de Bernoulli, X, *número de aciertos al realizar un tiro libre*, cuya distribución es:

$$X : \left| \begin{array}{ll} 1 & p[X=1] = p(\text{canasta}) = 0.8 \\ 0 & p[X=0] = p(\text{no canasta}) = 0.2 \end{array} \right. \triangleleft$$

9.3 Distribución binomial

Supongamos que se realizan n experimentos independientes de Bernoulli (usualmente n repeticiones independientes del mismo experimento), y que, en todos ellos, la probabilidad de éxito, p, es la misma. Se dice, entonces, que hemos realizado n *tiradas*[2] de Bernoulli de parámetro p. Sea X la variable *número de éxitos obtenidos*

2. La palabra *tiradas* recuerda a los juegos de azar, que son los experimentos aleatorios para los que primero se introdujo esta variable.

en las n tiradas. La distribución de la variable X se denomina **distribución binomial** de parámetros n y p. El parámetro n es el número de tiradas (un número entero positivo), y el parámetro p la probabilidad de éxito en una tirada (un valor entre cero y la unidad, esto es, $0 \leq p \leq 1$).

El hecho de que X siga una distribución binomial con parámetros n y p, lo denotaremos abreviadamente como

$$X \hookrightarrow \mathcal{B}(n,p) \,.$$

La variable aleatoria binomial fue presentada y obtenida por J. BERNOULLI en 1713, en su obra *Ars Conjectandi*, aunque los coeficientes binomiales que, como veremos, intervienen en su función de probabilidad, se conocían desde mucho antes. Es, pues, una de las variables aleatorias más antiguas.

Hallemos su función de probabilidad. Los valores de la variable son $k = 0, 1, \ldots, n$. Calculemos la probabilidad de obtener k éxitos, esto es, $p[X = k]$. Uno de los sucesos para el que tendremos k éxitos es el que resulta de obtener éxito en las primeras k tiradas y fracaso en las $n - k$ restantes, o sea

$$A \overset{k}{\ldots} A \overline{A} \overset{n-k}{\ldots} \overline{A} \,.$$

La probabilidad de este suceso, dada la independencia de los experimentos, es

$$\begin{aligned} p(A \overset{k}{\ldots} A \overline{A} \overset{n-k}{\ldots} \overline{A}) &= p(A) \overset{k}{\ldots} p(A) \cdot p(\overline{A}) \overset{n-k}{\ldots} p(\overline{A}) = \\ &= p^k \cdot (1-p)^{n-k} \,. \end{aligned}$$

Por otra parte, si contamos el número de sucesos en los que A aparezca k veces y $\overline{A}$ aparezca $n - k$ veces, nos encontraremos con que esta cantidad es el número combinatorio

$$PR_n^{k,n-k} = \frac{n!}{k! \cdot (n-k)!} = \binom{n}{k} \,.$$

Para todos estos sucesos, la probabilidad es la misma, es decir, $p^k \cdot (1-p)^{n-k}$. Por tanto, la **función de probabilidad** de una distribución binomial de parámetros n y p es

$$f_X(k) = p[X = k] = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \,, \qquad \text{para } k = 0, 1, \ldots, n \,.$$

Como puede observarse, si $p = 0$ la variable tiene una distribución degenerada en 0, esto es,

$$p[X = 0] = 1$$

y

$$p[X = 1] = \cdots = p[X = n] = 0 \,.$$

Recíprocamente, si $p = 1$, la distribución de la variable es degenerada en n, es decir,

$$p[X = n] = 1$$

y

$$p[X = 0] = \cdots = p[X = n-1] = 0 \,.$$

Para valores de p entre cero y la unidad, la probabilidad de un valor entero k cualquiera entre 0 y n es distinta de cero.

Obsérvese que las variables con distribución de Bernoulli de parámetro p, $b(p)$, pueden considerarse con distribución binomial de parámetros $n = 1$ y p, es decir, $\mathcal{B}(1, p)$.

Comprobemos que la función de probabilidad que hemos construido está bien definida. Evidentemente, todas las probabilidades son mayores o iguales que cero y, además, la suma de todas ellas es:

$$\sum_{k=0}^{n} p[X = k] = \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = [p + (1-p)]^n = 1^n = 1 .$$

Por lo tanto, la función de probabilidad está bien definida, de acuerdo con la proposición 4.8.

En la figura 9.3 se representa la función de probabilidad de una distribución binomial, en concreto de una binomial $\mathcal{B}(4, 0.3)$.

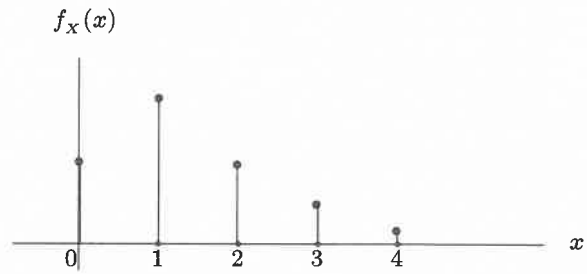


FIG. 9.3.

Un ejemplo de variable con distribución binomial es el número de caras obtenido al lanzar una moneda 100 veces, siempre que se garantice que los lanzamientos son independientes y que los sucesivos lanzamientos no modifican la probabilidad de obtener cara en cada uno de los siguientes.

Ejemplo 9.2. En el contexto del ejemplo 9.1, supongamos que el jugador realiza una serie de 5 tiros libres. ¿Cuál es la probabilidad de que enceste al menos dos?

La variable X, *número de aciertos en los 5 lanzamientos*, es una binomial de parámetros $n = 5$ y $p = 0.8$, con lo cual

$$p[X \geq 2] = \sum_{k=2}^{5} p[X = k] = \sum_{k=2}^{5} \binom{5}{k} 0.8^k \cdot 0.2^{5-k} = 0.99328 .$$

El lector aligerará los cálculos si tiene en consideración el suceso complementario, es decir,

$$p[X \geq 2] = 1 - p[X \leq 1] = 1 - \sum_{k=0}^{1} \binom{5}{k} 0.8^k \cdot 0.2^{5-k} \ . \triangleleft$$

El comportamiento de la función de probabilidad de una distribución binomial queda explicado por la proposición siguiente.

Proposición 9.1 *Sea X una variable con distribución binomial $\mathcal{B}(n,p)$. Si $(n+1)p$ no es entero, la función de probabilidad de X es creciente para $k < (n+1)p$ y decreciente para $k > (n+1)p$.*

Si $(n+1)p$ es un entero, $m = (n+1)p$, dicha función de probabilidad es creciente para $k \leq m-1$ y decreciente para $k \geq m$. Además, $p[X = m-1] = p[X = m]$.

Demostración. Calculemos el cociente entre dos valores consecutivos de la función de probabilidad.

$$\frac{p[X=k]}{p[X=k-1]} = \frac{\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}}{\binom{n}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-(k-1)}} = \frac{(n-k+1)p}{k(1-p)} \ .$$

Por tanto, $p[X=k]$ será mayor que $p[X=k-1]$ cuando

$$(n-k+1)p > k(1-p)$$

y, consecuentemente, cuando k sea menor que $(n+1)p$.

Igualmente, $p[X=k]$ será menor que $p[X=k-1]$, si k es mayor que $(n+1)p$.

Si $(n+1)p$ es un número entero (lo que depende del valor de p), la función de probabilidad toma el mismo valor para $k = (n+1)p$ y para el valor anterior, $k-1 = (n+1)p-1$, ya que el cociente de los dos valores de la función de probabilidad es igual a la unidad. ♣

Puede calcularse la moda de una variable aleatoria binomial utilizando la proposición anterior. Si $(n+1)p$ es un número entero, m, la distribución será bimodal, siendo $(m-1)$ y m las dos modas. Si $(n+1)p$ no es entero, la moda será el mayor entero inferior a $(n+1)p$, esto es, la parte entera de dicho número.

La expresión de la **función de distribución** de esta variable es:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{k} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} & k \leq x < k+1 \quad k = 0, \ldots, n-1 \\ 1 & x \geq n \end{cases}$$

La siguiente proposición indica que la variable binomial, $\mathcal{B}(n,p)$, podría obtenerse sumando n variables de Bernoulli de parámetro p e independientes, que representarían el resultado de cada tirada.

Proposición 9.2 *La variable aleatoria con distribución binomial es una suma de variables de Bernoulli de parámetro p e independientes. Es decir, si* $X \hookrightarrow \mathcal{B}(n,p)$, *entonces,*

$$X = X_1 + \cdots + X_n \ ,$$

donde $X_1, \ldots, X_n$ *son variables con distribución Bernoulli de parámetro p e independientes.*

La demostración es inmediata, sin más que utilizar la definición de una variable aleatoria binomial.

El cálculo de los momentos y la función característica resulta sencillo a partir de la proposición anterior, pues conocemos los momentos y la función característica de la distribución de Bernoulli, y tenemos expresada la variable binomial, X, como suma de n variables de Bernoulli, $X_1, \ldots, X_n$, independientes: $X = X_1 + \cdots + X_n$.

Esperanza:

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = np \ .$$

Varianza:

$$Var(X) = Var(X_1 + \cdots + X_n)$$

y, por ser $X_1, \ldots, X_n$ independientes, deducimos que

$$Var(X) = Var(X_1) + \cdots + Var(X_n) = npq \ .$$

Función característica:

Utilizando el hecho de que las n variables en que se descompone X son independientes y todas ellas tienen la misma distribución $b(p)$, cuya función característica ya conocemos, resulta:

$$\varphi_X(t) = \varphi_{\sum_{j=1}^{n} X_j}(t) = [\varphi_{X_j}(t)]^n = (q + p \cdot e^{it})^n \ .$$

Como vimos en el capítulo 8, el estudio de las sumas de variables aleatorias independientes va a ocupar un papel importante en la Inferencia Estadística. De especial interés resultan las variables tales que, si sumamos varias de ellas, la suma tiene una distribución del mismo tipo (aunque, tal vez, con distintos parámetros). Se dice que esas variables, y sus distribuciones de probabilidad, son **reproductivas para la suma**, o más brevemente, reproductivas, porque reproducen, mediante la suma, el mismo tipo de variable. Veamos que la distribución binomial es reproductiva.

Proposición 9.3 *Sean* $X_1, \ldots, X_m$, *m variables binomiales independientes con el mismo parámetro p,*

$$X_j \hookrightarrow \mathcal{B}(n_j, p) \ , \qquad para\ j = 1, \ldots, m \ .$$

Entonces, la variable $X = X_1 + \cdots X_m$ *sigue una distribución binomial,*

$$X \hookrightarrow \mathcal{B}(n_1 + \cdots + n_m, p) \ .$$

Demostración. Calculemos la función característica de X:

$$\varphi_X(t) = \varphi_{\sum_{j=1}^m X_j}(t) = \prod_{j=1}^m \varphi_{X_j}(t) = \left(q + p \cdot e^{it}\right)^{n_1+\cdots+n_m}$$

Obviamente, se trata de la función característica de una distribución binomial de parámetros $n_1 + \cdots + n_m$ y p, es decir, $\mathcal{B}(n_1 + \cdots + n_m, p)$. Por tanto, teniendo en cuenta el teorema de unicidad (corolario 8.6.1), la distribución de probabilidades de la suma, $X = X_1 + \cdots + X_n$, es una $\mathcal{B}(n_1 + \cdots + n_m, p)$. ♣

La siguiente proposición ayudará a calcular probabilidades de sucesos referentes a una variable binomial. En ella se establece la relación existente entre una variable $\mathcal{B}(n,p)$ y una $\mathcal{B}(n, 1-p)$.

Proposición 9.4 *Si* $X \hookrightarrow \mathcal{B}(n,p)$ *e* $Y \hookrightarrow \mathcal{B}(n, 1-p)$, *entonces,*

$$p[X = k] = p[Y = n - k] .$$

Demostración. Es evidente, puesto que

$$\begin{aligned} p[X = k] &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \binom{n}{n-k} \cdot (1-p)^{n-k} \cdot p^k = \\ &= p[Y = n - k] . \end{aligned}$$

El lector puede intuir el resultado: la probabilidad de obtener k éxitos en n tiradas coincide, lógicamente, con la probabilidad de obtener $n - k$ fracasos (de hecho, se trata del mismo suceso contabilizado de forma distinta). ♣

En el anexo final de este libro encontrará el lector una tabla que suministra los valores de la función de distribución de una variable binomial $\mathcal{B}(n,p)$ para $n \leq 20$ y $p \leq 0.5$. Para $p > 0.5$, se utiliza la proposición anterior, y si $n > 20$, existen aproximaciones de los valores de la función de distribución que describiremos más adelante.

9.4 Distribución de Poisson

Diremos que una variable aleatoria sigue una **distribución de Poisson** de parámetro λ ($\lambda > 0$), si su **función de probabilidad** es

$$f_X(k) = p[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} , \quad \text{para } k = 0, 1, \ldots$$

La distribución de Poisson tiene, por tanto, un único parámetro, λ, que es un número positivo y que, como veremos más adelante, coincide con la esperanza y con la varianza de la distribución.

Si una variable sigue esta distribución lo denotaremos por

$$X \hookrightarrow \mathcal{P}(\lambda) .$$

Comprobemos que la función de probabilidad está bien definida. En efecto, las probabilidades son positivas y su suma vale[3]

$$\sum_{k=0}^{\infty} p[X=k] = \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = e^0 = 1 .$$

En la figura 9.4 se representa la función de probabilidad de una distribución de Poisson, en concreto, de una distribución $\mathcal{P}(1.5)$.

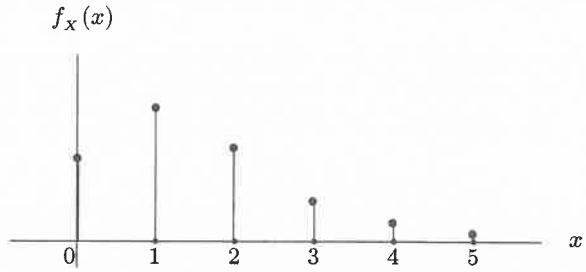


FIG. 9.4.

Aunque una variable de Poisson toma infinitos valores, sólo consideramos los seis primeros, pues, en este caso, ellos acumulan el 99.55 % de la probabilidad.

El estudio del crecimiento y decrecimiento de la función de probabilidad de una variable aleatoria de Poisson, y la obtención de la moda (o las modas) de esta distribución, sigue un razonamiento análogo al utilizado para la distribución binomial (proposición 9.1), quedando como ejercicio propuesto para el lector.

Históricamente, la distribución de Poisson aparece como límite de la distribución binomial cuando el número de experimentos tiende a infinito y la probabilidad de éxito tiende hacia cero. Más en concreto, fue propuesta por dicho autor en 1837 como límite de distribuciones binomiales, cuando el número de tiradas, n, tiende a infinito y la probabilidad de éxito vale $p = \frac{\lambda}{n}$, siendo λ un número positivo (por tanto, p tiende a cero, aunque $\lambda = n \cdot p$ puede ser de tamaño apreciable).

Aparece, entonces, cuando se estudian n tiradas de Bernoulli independientes con la misma probabilidad de éxito, p, con la condición de que n sea grande y p pequeña. Esta idea es la que hace que la distribución de Poisson se denomine *de los sucesos raros* (aludiendo al tamaño de p) o, en palabras de L. VON BORTKIEWICZ, la *ley de los números pequeños*. Este autor utilizó dicha variable como límite de la binomial,

3. Recordemos la expresión del desarrollo en serie de potencias de e^{λ},

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} .$$

estudiando (en 1898) el número de soldados del ejército prusiano muertos al año por coces de mulos. Como puede observarse, el número, n, de soldados expuestos a la acción de los animales es grande (o lo era en su día) y la probabilidad de que un soldado concreto fuese coceado puede aceptarse que es pequeña (parece difícil de aceptar, con todo, que dicha probabilidad sea la misma para toda la tropa).

W. S. GOSSET, *Student*, la utilizó en 1907 para estudiar el número de partículas que chocan contra una pequeña fracción de una amplia superficie expuesta a una radiación.

La siguiente proposición confirma los anteriores comentarios, es decir, que la función de probabilidad de una distribución de Poisson es el límite de la de una binomial, bajo ciertas condiciones.

Proposición 9.5 *Sea X una variable aleatoria con distribución binomial $\mathcal{B}(n,p)$. Si $n \to \infty$, $p \to 0$ y $\lambda = n \cdot p$ permanece constante, su función de probabilidad tiende hacia la de una distribución de Poisson de parámetro λ, es decir,*

$$\lim_{\substack{n\to\infty \\ p\to 0\ ,\ np=\lambda}} p[X=k] = p[Y=k]\ ,$$

siendo Y una variable aleatoria con distribución de Poisson de parámetro λ.

Demostración.

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ p\to 0\ ,\ np=\lambda}} p[X=k] &= \lim_{\substack{n\to\infty \\ p\to 0\ ,\ np=\lambda}} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \\
&= \lim_{n\to\infty} \frac{n\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \\
&= \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \frac{n\cdots(n-k+1)}{n^k} \cdot \frac{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n}{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^k} = \\
&= \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \left[\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}}\right] = \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \\
&= p[Y=k]\ ,
\end{aligned}$$

donde $Y \hookrightarrow \mathcal{P}(\lambda)$, sin más que recordar que

$$\lim_{x\to\infty} \left(1-\frac{1}{x}\right)^x = e^{-1}\ . \clubsuit$$

En el capítulo 13 (proposición 13.3), se demuestra la extensión de esta proposición a funciones de distribución. En la práctica, y en virtud de este resultado, se puede considerar razonable la aproximación de la distribución binomial a la de Poisson, si $n > 20$ y $p \leq 0.1$.

La distribución de Poisson se asemeja, por tanto, a una binomial con probabilidad de éxito muy pequeña (es decir, el éxito es un suceso *raro*) y un número grande de tiradas. Las situaciones en las que se utiliza esta distribución siguen habitualmente este esquema; así, se ajustarán a una distribución de Poisson variables como

el número de accidentes laborales diarios de una gran fábrica, el número de erratas en una página de un libro, el de llamadas a una centralita telefónica, el número de personas que cumplen años el día de Navidad, etc.

Su **función de distribución** es:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!} & k \leq x < k+1 \quad k = 0, 1, \ldots \end{cases}$$

El cálculo de sus momentos y de su función característica es sencillo, aunque laborioso.

Esperanza:

$$E(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda \, .$$

Varianza:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} [(k-1)+1] \cdot \frac{\lambda^{k-1}}{(k-1)!} = \\ &= \lambda \cdot e^{-\lambda} \cdot \left[\lambda \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] = \\ &= \lambda \cdot e^{-\lambda} \cdot (\lambda \cdot e^{\lambda} + e^{\lambda}) = \lambda \cdot (\lambda + 1) = \lambda^2 + \lambda \, . \end{aligned}$$

Por tanto,

$$Var(X) = E(X^2) - [E(X)]^2 = \lambda \, .$$

Función característica[4]:

$$\begin{aligned} \varphi_X(t) &= E(e^{itX}) = \sum_{k=0}^{\infty} e^{itk} \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!} = \\ &= e^{-\lambda} \cdot e^{e^{it}\lambda} = e^{\lambda(e^{it}-1)} \, . \end{aligned}$$

Puede el lector obtener los momentos, derivando esta función característica, y comprobar que coinciden con los que hemos calculado directamente.

Ejemplo 9.3. Nuestro experto jugador de baloncesto tiene una media de un enceste por temporada desde una distancia de diez metros. Suponiendo que el número de canastas en lanzamientos de este tipo sigue una distribución de Poisson, ¿cuál es la probabilidad de que haga dos encestes desde dicha distancia en la próxima temporada?

4. Acepte el lector que las exponenciales complejas se desarrollan en serie como las reales, esto es, si z es un número complejo,

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \, .$$

Denotando por X la variable *número de canastas en tiros desde 10 metros*, tendremos que calcular $p[X = 2]$. Considerando que la distribución de esta variable es una distribución de Poisson de parámetro 1 resultará:

$$p[X = 2] = e^{-1} \cdot \frac{1^2}{2!} = 0.1839 \,. \triangleleft$$

La siguiente proposición muestra que la distribución de Poisson es reproductiva para cualesquiera valores del parámetro.

Proposición 9.6 *La suma de variables de Poisson independientes es una variable de Poisson.*

Demostración. Sean m variables de Poisson independientes de parámetro λ_j, $j = 1, \ldots, m$,

$$X_j \hookrightarrow \mathcal{P}(\lambda_j) \,.$$

Sea X su suma, es decir,

$$X = X_1 + \cdots + X_m \,.$$

Entonces, calculando la función característica de ésta, tendremos

$$\begin{aligned} \varphi_X(t) &= \varphi_{\sum_{j=1}^m X_j}(t) = \prod_{j=1}^m \varphi_{X_j}(t) = \\ &= e^{\lambda_1(e^{it}-1)} \cdots e^{\lambda_m(e^{it}-1)} = e^{\left(\sum_1^m \lambda_j\right)(e^{it}-1)} \,, \end{aligned}$$

que es la función característica de una variable con distribución de Poisson de parámetro $\sum_{j=1}^m \lambda_j$. En consecuencia, utilizando el teorema de unicidad, X tiene una distribución de Poisson de parámetro $\sum_{j=1}^m \lambda_j$. ♣

La siguiente proposición establece una curiosa relación adicional entre las distribuciones de Poisson y binomial.

Proposición 9.7 *Sean X e Y dos variables aleatorias independientes e igualmente distribuidas de Poisson de parámetro λ. La distribución de X condicionada por $X + Y = n$ es una binomial de parámetros n y $\frac{1}{2}$.*

Demostración. En efecto, $X + Y \hookrightarrow \mathcal{P}(2\lambda)$ por la proposición anterior y, por otra parte, la probabilidad condicionada es:

$$\begin{aligned} p[X = k \mid X + Y = n] &= \frac{p[X = k, Y = n - k]}{p[X + Y = n]} = \frac{p[X = k] \cdot p[Y = n - k]}{p[X + Y = n]} = \\ &= \frac{e^{-\lambda}\lambda^k/k! \cdot e^{-\lambda}\lambda^{n-k}/(n-k)!}{e^{-2\lambda}(2\lambda)^n/n!} = \binom{n}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{n-k} \,, \end{aligned}$$

que es la función de probabilidad de una binomial $\mathcal{B}\left(n, \frac{1}{2}\right)$. ♣

El anexo final del libro proporciona los valores de la función de distribución de una variable de Poisson de parámetro $\lambda \leq 20$. Para mayores valores del parámetro, se utilizan aproximaciones a otras distribuciones.

9.5 Distribución geométrica

Supongamos que realizamos sucesivos experimentos independientes de Bernoulli, con probabilidad fija de éxito, $p(A) = p$. Definimos la variable X como el *número de fracasos antes del primer éxito*; se dice que X sigue una **distribución geométrica** de parámetro p $(0 < p \leq 1)$. Dicha distribución es conocida, también, como distribución de Pascal.

Teniendo en cuenta que la probabilidad de obtener k fracasos antes del primer éxito es

$$p(\overline{A}\ldots\overset{k}{\ldots}\ldots\overline{A}A) = \left[p(\overline{A})\right]^k \cdot p(A)\ ,$$

la **función de probabilidad** de esta variable será

$$f_X(k) = p[X = k] = q^k \cdot p\ , \qquad \text{para } k = 0, 1, \ldots\ ,$$

donde, como antes, $q = 1 - p$.

El parámetro de la distribución es la probabilidad, p, de obtener éxito en una tirada. Por motivos obvios, se descarta que p pueda ser igual a cero, ya que, entonces, no se obtendría nunca un éxito y la variable no estaría definida. Si p es igual a la unidad, q es igual a cero y la distribución es degenerada (la variable vale siempre cero), ya que se obtendría necesariamente un éxito en la primera tirada; en concreto, la función de probabilidad vale

$$p[X = 0] = 0^0 \cdot 1 = 1$$

y

$$p[X = k] = 0^k \cdot 1 = 0\ ,\ \text{si } k > 0\ .$$

El hecho de que una variable aleatoria X siga una distribución geométrica se expresa abreviadamente por $X \hookrightarrow \mathcal{G}(p)$.

Es inmediato comprobar que la función de probabilidad está bien definida. Evidentemente es no negativa y, calculando la suma de sus probabilidades[5],

$$\sum_{k=0}^{\infty} p[X = k] = \sum_{k=0}^{\infty} q^k p = p \cdot \sum_{k=0}^{\infty} q^k = p \cdot \frac{1}{1-q} = 1\ .$$

En la figura 9.5 se representa la función de probabilidad de una distribución geométrica.

5. Recordemos que la suma de una progresión geométrica ilimitada con razón $|r| < 1$ es

$$S = \frac{a_1}{1-r}\ ,$$

donde a_1 es el primer término de la progresión.

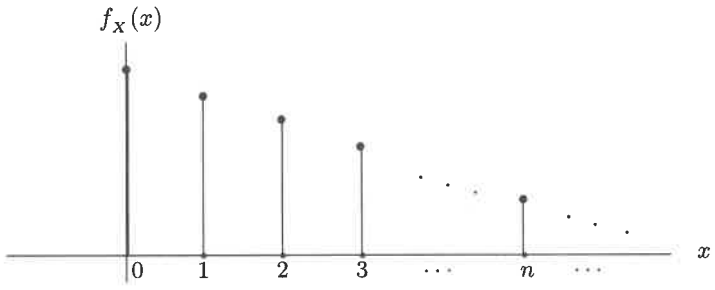


FIG. 9.5.

Como la serie geométrica con razón menor que la unidad es siempre decreciente, también lo será la función de probabilidad de una distribución geométrica. Por tanto, la moda de una variable geométrica es siempre el menor valor de la misma, es decir, $Mo = 0$.

Ejemplo de este tipo de variable es el número de tiradas de un dado antes de conseguir un seis.

Esta distribución se utiliza en auditoría, cuando se inspecciona, por ejemplo, un lote de facturas con el fin de estimar la proporción, p, que entre ellas corresponden a un cierto proveedor, extrayendo al azar facturas del lote y observando el número de extracciones precisas hasta que aparezca la primera correspondiente a dicho proveedor. En este caso, cada extracción es una tirada, y el suceso que consideramos como éxito es que la factura extraída sea del proveedor en cuestión. La probabilidad de éxito será, obviamente, la proporción de esas facturas en el lote, y dicha probabilidad se mantendrá fija, si las facturas inspeccionadas se devuelven al lote tras cada extracción.

La **función de distribución** de una variable geométrica es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{k} q^i \cdot p = 1 - q^{k+1} & k \leq x < k+1 \quad \text{para } k = 0, 1, \ldots \end{cases}$$

Esta función de distribución es escalonada, siendo sus *saltos* cada vez más pequeños y no alcanzándose nunca la unidad.

Los momentos de esta distribución se obtienen más fácilmente si calculamos previamente la función característica.

Función característica[6]:

$$\varphi_X(t) = E(e^{itX}) = \sum_{k=0}^{\infty} e^{itk} \cdot q^k \cdot p = p \cdot \sum_{k=0}^{\infty} (e^{it}q)^k = \frac{p}{1 - e^{it} \cdot q} .$$

6. Debe aceptar el lector que, si $\sum_{k=0}^{\infty} z^k$ es una serie geométrica compleja con $|z| < 1$, su suma se calcula como en el caso real y, por tanto, vale

$$\sum_{k=0}^{\infty} z^k = \frac{z^0}{1-z} = \frac{1}{1-z}.$$

Hallemos, a partir de la función característica, los momentos, utilizando la proposición 8.4. Derivando la función característica,

$$\varphi'_X(t) = \frac{ipqe^{it}}{(1 - e^{it} \cdot q)^2}$$

y derivando de nuevo

$$\varphi''_X(t) = \frac{i^2pq(e^{it} + e^{2it}q)}{(1 - e^{it} \cdot q)^3} .$$

Haciendo $t = 0$,

$$\varphi'_X(0) = \frac{ipq}{(1-q)^2} = \frac{iq}{p} ,$$
$$\varphi''_X(0) = \frac{i^2pq(1+q)}{(1-q)^3} = \frac{i^2q(1+q)}{p^2} .$$

Podemos deducir, entonces, los momentos de primer y segundo orden; la **esperanza** vale

$$E(X) = \frac{\varphi'_X(0)}{i} = \frac{q}{p} ,$$

y el momento no central de segundo orden,

$$E(X^2) = \frac{\varphi''_X(0)}{i^2} = \frac{q(1+q)}{p^2} .$$

En consecuencia, la **varianza** valdrá:

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{q}{p^2} .$$

Ejemplo 9.4. ¿Cuál es el número medio de fallos del jugador de baloncesto (ejemplo 9.1) antes de conseguir el primer acierto en tiro libre, sabiendo que acierta en el 80 % de las ocasiones?

Estamos interesados en la esperanza de la variable X, *número de tiros libres no encestados antes del primero encestado*, variable cuya distribución es geométrica de parámetro $p = 0.8$. Bastará sustituir en la expresión de la esperanza para obtener:

$$E(X) = \frac{0.2}{0.8} = 0.25 .\triangleleft$$

Una importante propiedad que cumplen estas distribuciones es la propiedad denominada del *olvido* o de la *falta de memoria*, que estudiamos a continuación.

Proposición 9.8 *Sea X una variable aleatoria con distribución $\mathcal{G}(p)$. Entonces,*

$$p[X \geq m + n \mid X \geq m] = p[X \geq n] , \qquad m, n = 0, 1, \ldots$$

Recíprocamente, si una variable aleatoria toma valores enteros no negativos y verifica la igualdad anterior, es necesariamente una variable aleatoria geométrica.

Demostración. En efecto, dada una variable geométrica, X,

$$\begin{aligned} p[X \geq k] &= 1 - p[X < k] = 1 - F_X(k-1) = \\ &= 1 - (1 - q^k) = q^k \ , \end{aligned}$$

para $k = 0, 1, \ldots$; por tanto,

$$\begin{aligned} p[X \geq m+n \mid X \geq m] &= \frac{p[X \geq m+n, X \geq m]}{p[X \geq m]} = \frac{p[X \geq m+n]}{p[X \geq m]} = \\ &= \frac{q^{m+n}}{q^m} = q^n = p[X \geq n] \ . \end{aligned}$$

Demostremos, ahora, la afirmación recíproca. De la propiedad del olvido se deduce que

$$\frac{p[X \geq m+n]}{p[X \geq m]} = p[X \geq n] \ ,$$

luego,

$$p[X \geq m+n] = p[X \geq m] \cdot p[X \geq n] \ .$$

Sustituyendo n por $n+1$, se tiene:

$$p[X \geq m+n+1] = p[X \geq m] \cdot p[X \geq n+1] \ .$$

Planteando la diferencia entre ambas expresiones, obtenemos:

$$\begin{aligned} p[X = m+n] &= p[X \geq m+n] - p[X \geq m+n+1] = \\ &= p[X \geq m] \cdot (p[X \geq n] - p[X \geq n+1]) = \\ &= p[X \geq m] \cdot p[X = n] \ . \end{aligned}$$

Si hacemos $m = 1$ y $n = 0$, resulta:

$$p[X = 1] = p[X \geq 1] \cdot p[X = 0] = (1 - p[X = 0]) \cdot p[X = 0] \ .$$

Ahora bien, como $p = p[X = 0]$ y $q = 1 - p$, entonces,

$$p[X = 1] = q \cdot p \ .$$

Si $m = 1$ y $n = 1$,

$$p[X = 2] = p[X \geq 1] \cdot p[X = 1] = q \cdot (q \cdot p) = q^2 \cdot p \ .$$

Actuando por recurrencia sobre n, resulta:

$$p[X = k] = q^k \cdot p \ , \qquad \text{para } k = 0, 1, 2, \ldots$$

Por tanto, la variable X sigue una distribución geométrica, siendo p el parámetro. ♣

La propiedad que demuestra la proposición anterior se denomina del *olvido* porque la igualdad de probabilidades que aparece en el enunciado puede describirse de la siguiente forma: si hemos rebasado la tirada m-ésima sin que se haya obtenido un éxito, la probabilidad de que transcurran otras n tiradas, al menos, sin que se presente (esto es, la probabilidad de que el número de tiradas antes del primer éxito sea, al menos, $m+n$), es la misma que se obtendría si nos olvidáramos de las

tiradas anteriores y consideráramos que la primera tirada es la $(m+1)$-ésima, es decir, la probabilidad de que transcurran al menos n tiradas sin obtener un éxito.

Esta propiedad da pie a que se asignen distribuciones geométricas a los experimentos aleatorios basados en la observación repetida de un fenómeno, cuando resulte aceptable que la historia de las observaciones no condiciona su devenir futuro, al menos si, hasta entonces, el resultado que se espera no se ha presentado. La segunda parte de la proposición anterior indica que la asignación de una distribución geométrica a una variable que marca el tiempo de espera (el que tarda en presentarse el resultado) de un proceso sin *memoria* es la más razonable, siempre que la variable sea discreta, esto es, siempre que el tiempo se mida a intervalos.

Como la función de distribución de una variable geométrica tiene una expresión fácilmente calculable, con la ayuda de una calculadora de bolsillo, no se precisa el uso de tablas para la obtención de sus probabilidades.

9.6 Distribución binomial negativa

Supongamos que se realizan sucesivos experimentos de Bernoulli independientes, donde la probabilidad de *éxito* es $p(A) = p$ (o sea, la misma situación que servía para definir la variable geométrica). La variable X, *número de fracasos antes del r-ésimo éxito*, se dice que sigue una **distribución binomial negativa** de parámetros r y p, con $r = 1, 2, \ldots$ y $0 < p \leq 1$.

Calculemos la probabilidad de obtener k fracasos antes del r-ésimo éxito. Uno de los sucesos para los que esto ocurre se obtiene cuando las primeras k tiradas sean fracasos y las r siguientes éxitos, es decir,

$$\overline{A} \overset{k}{\ldots} \overline{A} A \overset{r}{\ldots} A;$$

dada la independencia de las tiradas y, por tanto, de los sucesos, su probabilidad será

$$\left[p(\overline{A})\right]^k \cdot [p(A)]^r = q^k \cdot p^r .$$

Puesto que todos los sucesos que proporcionan k fracasos antes del r-ésimo éxito son igualmente probables, nos resta conocer cuántos hay de este tipo. Como el último de los resultados es necesariamente un éxito, el número buscado es

$$PR^{k,r-1}_{k+r-1} = \binom{k+r-1}{k} .$$

En consecuencia, la expresión de la **función de probabilidad** de la distribución binomial negativa de parámetros r y p es

$$f_X(k) = p[X = k] = \binom{k+r-1}{k} \cdot (1-p)^k \cdot p^r , \quad \text{para } k = 0, 1, \ldots$$

Esta distribución tiene dos parámetros; el primero, r, es el número de éxitos que deseamos encontrar y, consecuentemente, será un entero positivo ($r = 1, 2, \ldots$). Para $r = 1$ se tiene, obviamente, la distribución geométrica. El segundo parámetro,

p, es la probabilidad de éxito en una tirada; varía, por tanto, entre cero y la unidad, y no puede ser igual a cero porque nunca se obtendría un éxito (la variable no estaría definida). Si p es igual a la unidad, el resultado será éxito en todas las tiradas, por lo que la variable valdría siempre cero, esto es, sería degenerada. Esta última afirmación se puede comprobar, también, calculando la función de probabilidad para $p = 1$.

Si X tiene una distribución binomial negativa con parámetros r y p, se designa abreviadamente como

$$X \hookrightarrow \mathcal{BN}(r, p) .$$

Sin más que considerar la igualdad

$$\binom{k+r-1}{k} = (-1)^k \cdot \binom{-r}{k} ,$$

la función de probabilidad de esta variable puede escribirse, asimismo, como

$$p[X = k] = \binom{-r}{k} \cdot (-q)^k \cdot p^r , \qquad \text{para } k = 0, 1, \ldots$$

La comparación de esta expresión con la función de probabilidad de la distribución binomial justifica el nombre de distribución binomial negativa con el que se designa a la primera.

Comprobemos que es, en efecto, una función de probabilidad[7].

$$\sum_{k=0}^{\infty} p[X = k] = \sum_{k=0}^{\infty} \binom{-r}{k} \cdot (-q)^k \cdot p^r = p^r \cdot (-q+1)^{-r} = 1 .$$

La figura 9.6 representa la función de probabilidad de una binomial negativa, en concreto, la de una binomial negativa $\mathcal{BN}(3, 0.5)$.

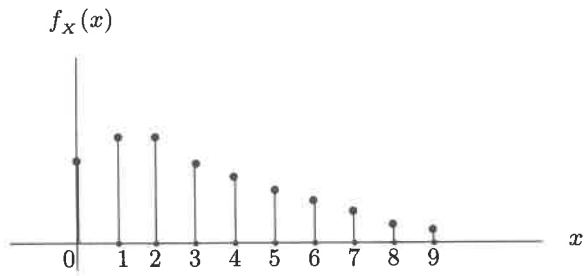


FIG. 9.6.

7. Recordemos que

$$(t+1)^{-n} = \sum_{k=0}^{\infty} \binom{-n}{k} \cdot t^k ,$$

siendo $\binom{-n}{k} = \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!}$.

La variable aleatoria binomial negativa aparece en un estudio de MONTMORT sobre los juegos de azar (1714), aunque sin formalizar ya se describe en unos estudios de B. PASCAL, 35 años antes.

Más adelante, se utiliza como alternativa a la distribución de Poisson, cuando ésta no es adecuada para describir el fenómeno. Así la propone W. S. GOSSET, *Student*, para estudiar el número de plaquetas en un hematocitómetro.

Asimismo, se emplea como distribución límite de ciertas variables más complicadas que aparecen en esquemas de contagios (modelos de urnas, básicamente).

Como ejemplo, podemos plantear el número de caras al lanzar una moneda, antes de obtener la octava cruz. Digamos, también, que al igual que la distribución geométrica, se utiliza en auditoría cuando se inspeccionan lotes de facturas y se espera la aparición de r de ellas correspondientes a un proveedor determinado.

Ejemplo 9.5. En una jornada de entrenamiento, el jugador de baloncesto realiza series de tiros libres. ¿Cuál es la probabilidad de que falle 5 canastas antes de conseguir 7 puntos?

En este caso, la variable X, *número de canastas erradas antes del séptimo acierto*, sigue una distribución binomial negativa con $p = 0.8$ y $r = 7$. La probabilidad que nos interesa es, por tanto,

$$p[X = 5] = \binom{5+7-1}{5} \cdot 0.2^5 \cdot 0.8^7 = 0.031 \; . \triangleleft$$

Escribamos la expresión de su **función de distribución**:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \sum_{i=0}^{k} \binom{i+r-1}{i} \cdot q^i \cdot p^r & k \leq x < k+1 \quad k = 0, 1, \ldots \end{cases}$$

La siguiente proposición muestra que la distribución binomial negativa es suma de variables geométricas independientes. Piense el lector que el número de fracasos antes del éxito r-ésimo resulta de sumar los recuentos parciales de fracasos entre dos éxitos consecutivos, hasta que se obtiene el r-ésimo.

Proposición 9.9 *Sea* $X \hookrightarrow \mathcal{BN}(r, p)$. *Entonces,*

$$X = X_1 + \cdots + X_r \; ,$$

donde $X_1, \ldots, X_r$ *son variables aleatorias independientes con distribución geométrica de parámetro* p.

Demostración. Consideremos las variables X_1, *número de fracasos antes del primer éxito*; X_2, *número de fracasos entre el primer y segundo éxito*, ...; X_r, *número de fracasos entre el* $(r-1)$*-ésimo y el* r*-ésimo éxito*. Obviamente, cada una de ellas sigue una distribución geométrica de parámetro p y, además,

$$X = X_1 + \cdots + X_r \; .$$

Finalmente, $X_1, \ldots, X_r$ son variables aleatorias independientes, ya que están asociadas a distintos experimentos y éstos son independientes. ♣

Adviértase que, así como las variables binomiales se podían entender como suma de variables de Bernoulli independientes, las binomiales negativas se pueden entender como suma de geométricas independientes. Por otro lado, tanto la distribución de Bernoulli como la geométrica son casos particulares de la binomial y de la binomial negativa, respectivamente; en concreto, $b(p) = \mathcal{B}(1,p)$ y $\mathcal{G}(p) = \mathcal{BN}(1,p)$.

La propiedad anterior facilita los cálculos de las características de una variable binomial negativa, X, al considerarla como suma de variables, X_i, geométricas e independientes: $X = X_1 + \cdots + X_n$.

Esperanza:

$$E(X) = E(X_1) + \cdots + E(X_r) = \frac{r \cdot q}{p} .$$

Varianza:

$$Var(X) = Var(X_1 + \cdots + X_r) = Var(X_1) + \cdots + Var(X_r) = \frac{r \cdot q}{p^2} .$$

Función característica:

$$\varphi_X(t) = \varphi_{X_1+\cdots+X_r}(t) = [\varphi_{X_i}(t)]^r = \left(\frac{p}{1 - e^{it} \cdot q}\right)^r .$$

La distribución binomial negativa es reproductiva, es decir, la suma de variables binomiales negativas de parámetro p, independientes, es una binomial negativa. El resultado se demuestra en la siguiente proposición.

Proposición 9.10 *Sean $X_1, \ldots, X_m$, m variables binomiales negativas independientes con el mismo parámetro p,*

$$X_j \hookrightarrow \mathcal{BN}(r_j, p) , \qquad con\ j = 1, \ldots, m .$$

Entonces, la variable $X = X_1 + \cdots + X_m$ sigue una distribución binomial negativa,

$$X \hookrightarrow \mathcal{BN}\left(\sum_{j=1}^{m} r_j,\ p\right) .$$

Demostración. En efecto, si $X_1, \ldots, X_m$ son binomiales negativas $\mathcal{BN}(r_j, p)$ $(j = 1, \ldots, m)$ e independientes, la función característica de $X = X_1 + \cdots + X_m$ es

$$\varphi_X(t) = \varphi_{\sum_{j=1}^m X_j}(t) = \prod_{j=1}^{m} \varphi_{X_j}(t) = \prod_{j=1}^{m} \left(\frac{p}{1 - e^{it} \cdot q}\right)^{r_j} = \left(\frac{p}{1 - e^{it} \cdot q}\right)^{\sum_{j=1}^m r_j} ,$$

que es la función característica de una $\mathcal{BN}\left(\sum_{j=1}^m r_j, p\right)$.

Por tanto, de acuerdo con el teorema de unicidad, la variable X tiene una distribución $\mathcal{BN}\left(\sum_{j=1}^m r_j, p\right)$. ♣

9.7 Distribución hipergeométrica

Sea una urna con N bolas ($N = 1, 2, \ldots$) de las cuales N_1 son blancas y N_2 son negras. Realizamos n extracciones sin reposición ($n \leq N$). Si X es el *número de bolas blancas obtenidas*, diremos que X sigue una **distribución hipergeométrica** de parámetros N, n y p, siendo $p = \frac{N_1}{N}$ la proporción inicial de bolas blancas en la urna (es decir, la probabilidad de obtener una bola blanca en la primera extracción). Una variable binomial negativa se designa abreviadamente como

$$X \hookrightarrow \mathcal{H}(N, n, p) .$$

Calculemos la probabilidad de obtener k bolas blancas al realizar n extracciones sin reposición.

Uno de los posibles sucesos es *obtener bola blanca en las k primeras extracciones y negra en las $n - k$ restantes*; es decir,

$$B_1 B_2 \ldots B_k \overline{B}_{k+1} \overline{B}_{k+2} \ldots \overline{B}_n ,$$

donde denotamos por B_i al suceso *obtener bola blanca en la i-ésima extracción* y por $\overline{B}_i$ *obtener negra en la misma.*

Teniendo en cuenta que las extracciones se hacen sin devolución y, por tanto, son dependientes, su probabilidad $p(B_1 B_2 \ldots B_k \overline{B}_{k+1} \overline{B}_{k+2} \ldots \overline{B}_n)$ será

$$\begin{aligned} p(B_1) \cdot P(B_2 \mid B_1) \ldots p(B_k \mid B_1 \ldots B_{k-1}) \cdot p(\overline{B}_{k+1} \mid B_1 \ldots B_k) \cdot \\ \cdot p(\overline{B}_{k+2} \mid B_1 \ldots B_k \overline{B}_{k+1}) \ldots p(\overline{B}_n \mid B_1 \ldots \overline{B}_{n-1}) . \end{aligned}$$

Es decir,

$$\frac{N_1}{N} \cdot \frac{N_1 - 1}{N - 1} \cdots \frac{N_1 - k + 1}{N - k + 1} \cdot \frac{N_2}{N - k} \cdot \frac{N_2 - 1}{N - k - 1} \cdots \frac{N_2 - n + k + 1}{N - n + 1} ,$$

o lo que es lo mismo,

$$\frac{\frac{N_1!}{(N_1-k)!} \cdot \frac{N_2!}{(N_2-n+k)!}}{\frac{N!}{(N-n)!}} .$$

Ahora bien, cualquier suceso en el que se obtengan k bolas blancas y $n - k$ negras, tiene la misma probabilidad. Contabilicemos el número de los mismos:

$$PR_n^{k,n-k} = \frac{n!}{k! \cdot (n-k)!} .$$

Por tanto,

$$p[X = k] = \frac{n!}{k! \cdot (n-k)!} \cdot \frac{\frac{N_1!}{(N_1-k)!} \cdot \frac{N_2!}{(N_2-n+k)!}}{\frac{N!}{(N-n)!}} = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}$$

es la **función de probabilidad** de la distribución hipergeométrica, que también podemos escribir como

$$f_X(k) = p[X = k] = \frac{\binom{N \cdot p}{k}\binom{N \cdot q}{n-k}}{\binom{N}{n}} ,$$

donde $p = \frac{N_1}{N}$ y $q = \frac{N_2}{N}$. El lector puede obtener esta función de probabilidad utilizando la regla de Laplace y siguiendo un razonamiento análogo al empleado en el ejemplo 3.13.

Por otra parte, y puesto que hay N_1 bolas blancas y N_2 negras, $k \leq \min(n, N_1)$ y $n - k \leq \min(n, N_2)$, con lo cual los valores posibles, k, de la variable, son los que verifican las desigualdades

$$\max(0, n - N_2) \leq k \leq \min(n, N_1) .$$

Veamos que la función de probabilidad está bien definida. Evidentemente, es no negativa y, además[8], sumando en k desde $\max(0, n - N_2)$ hasta $\min(n, N_1)$,

$$\sum_k p[X = k] = \sum_k \frac{\binom{N \cdot p}{k} \cdot \binom{N \cdot q}{n-k}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1 .$$

Un ejemplo de esta variable es el número de unidades defectuosas al hacer una inspección de un lote de productos, cuando la inspección destruye las unidades y, por tanto, no pueden reintegrarse al lote tras la misma. También, cuando se realiza una encuesta para estudiar si los individuos de una población tienen o no una determinada característica, el número de individuos de la muestra que poseen dicha característica sigue una distribución hipergeométrica; N sería el tamaño de la población, n el número de encuestados y p la proporción de individuos de la población que poseen la mencionada característica (que, habitualmente, es la cantidad que se desea conocer).

La expresión de su **función de distribución** es

$$F_X(x) = \begin{cases} 0 & x < \max(0, n - N_2) \\ \sum_{i=0}^{k} \frac{\binom{N \cdot p}{i} \cdot \binom{N \cdot q}{n-i}}{\binom{N}{n}} & k \leq x < k + 1 \\ 1 & x \geq \min(n, N_1) \end{cases}$$

donde los valores, k, que aparecen en la segunda de las expresiones, son aquellos enteros que cumplen $\max(0, n - N_2) \leq k < \min(n, N_1)$.

Ejemplo 9.6. El equipo de baloncesto cuenta con una plantilla de 8 jugadores para jugar en el puesto de alero, de los cuales 2 son extranjeros. El entrenador selecciona al

8. Recordemos dos importantes propiedades de los números combinatorios que utilizaremos en este apartado:

1. Para $l \geq 1$ se verifica que $\binom{m}{l} = \frac{m}{l} \cdot \binom{m-1}{l-1}$.
2. $\sum_{l=0}^{m} \binom{a}{l} \cdot \binom{b}{m-l} = \binom{a+b}{m}$, admitiendo que, si $l > s$, entonces $\binom{s}{l} = 0$.

azar dos de ellos para formar parte del cinco inicial. ¿Cuál es la probabilidad de que uno sea extranjero?

En este caso $N = 8$, $N_1 = 2$ y $n = 2$, con lo cual, si X es el *número de extranjeros elegidos*, la probabilidad que nos interesa es

$$p[X = 1] = \frac{\binom{2}{1}\binom{6}{1}}{\binom{8}{2}} = 0.428 \ .\triangleleft$$

Vamos a analizar, en un ejemplo, el papel que desempeña la distribución hipergeométrica en los llamados procedimientos de inspección por muestreo o muestreo de aceptación utilizados en la industria.

Ejemplo 9.7. En términos generales, la inspección por muestreo consiste en evaluar si una determinada colección de unidades es o no aceptable, a partir de la elección aleatoria de un subconjunto o muestra de la misma. Las unidades se clasifican en defectuosas y no defectuosas, y se presentan en lotes de N unidades. Denotaremos por d el número de defectuosas en un lote. Para inspeccionar un lote, se elige al azar un muestra de n unidades (distintas); sea X el *número de defectuosas en la muestra.*

El criterio utilizado en este tipo de procedimientos consiste en aceptar el lote si $X \leq c$, y rechazarlo en caso contrario. Es obvio que este criterio puede conducirnos a la aceptación de lotes con muchas unidades defectuosas y viceversa, es decir, al rechazo de lotes con pocas unidades defectuosas. Por ello, resulta de interés el cálculo de la probabilidad de aceptar un lote, suponiendo que contiene d unidades defectuosas, cálculo que resulta sencillo si tenemos en cuenta que X sigue una distribución hipergeométrica de parámetros N, n y $p = \frac{d}{N}$. Por tanto,

$$p[X = k \mid d \ \text{ defectuosas}] = \frac{\binom{d}{k}\binom{N-d}{n-k}}{\binom{N}{n}} \ ,$$

con lo cual la probabilidad de aceptación de un lote con d defectuosas es

$$p[X \leq c \mid d \ \text{ defectuosas}] = \sum_{k=0}^{c} \frac{\binom{d}{k}\binom{N-d}{n-k}}{\binom{N}{n}} \ .\triangleleft$$

Finalmente, calculemos la esperanza y la varianza de la distribución hipergeométrica, utilizando los resultados que aparecen en la nota anterior a pie de página.

Esperanza:

$$\begin{aligned} E(X) &= \sum_k k \cdot p[X = k] = \sum_k k \cdot \frac{\binom{N \cdot p}{k} \cdot \binom{N \cdot q}{n-k}}{\binom{N}{n}} = \\ &= \frac{N \cdot p}{\binom{N}{n}} \cdot \sum_k \binom{N \cdot p - 1}{k - 1} \cdot \binom{N \cdot q}{n - k} = \frac{N \cdot p}{\binom{N}{n}} \cdot \binom{N - 1}{n - 1} = n \cdot p \ . \end{aligned}$$

Varianza:

$$E(X^2) = \sum_k k^2 \cdot p[X=k] = \sum_k k^2 \cdot \frac{\binom{N\cdot p}{k}\cdot\binom{N\cdot q}{n-k}}{\binom{N}{n}} =$$
$$= \frac{N\cdot p}{\binom{N}{n}}\left[\sum_k [(k-1)+1]\binom{N\cdot p-1}{k-1}\cdot\binom{N\cdot q}{n-k}\right],$$

expresión que, sin más que hacer operaciones, es igual a

$$\frac{N\cdot p}{\binom{N}{n}}\left[(N\cdot p-1)\sum_k\binom{N\cdot p-2}{k-2}\cdot\binom{N\cdot q}{n-k}+\sum_k\binom{N\cdot p-1}{k-1}\cdot\binom{N\cdot q}{n-k}\right],$$

con lo cual,

$$E(X^2) = \frac{N\cdot p}{\binom{N}{n}}\cdot\left[(N\cdot p-1)\binom{N-2}{n-2}+\binom{N-1}{n-1}\right] =$$
$$= \frac{N\cdot p\cdot n!\cdot(N-n)!}{N!}\left[\frac{(N\cdot p-1)(n-1)(N-2)!+(N-1)!}{(n-1)!(N-n)!}\right] =$$
$$= \frac{n\cdot p\cdot[(N\cdot p-1)(n-1)+(N-1)]}{N-1},$$

obteniéndose, así, la varianza:

$$Var(X) = E(X^2) - [E(X)]^2 = npq\cdot\frac{N-n}{N-1}.$$

La diferencia entre la distribución hipergeométrica y la binomial radica en que en la binomial la probabilidad de éxito permanece constante en los n experimentos, mientras que en la hipergeométrica dicha probabilidad varía de un experimento a otro. En el esquema de urnas con bolas blancas y negras, la variable X, *número de bolas blancas obtenidas* al realizar n extracciones, tiene una distribución binomial si las extracciones se realizan con reemplazamiento, y una distribución hipergeométrica si se realizan sin reemplazamiento. Ahora bien, si el número de bolas es mucho mayor que el número de extracciones, cada extracción sin reposición no cambiará sustancialmente la composición de la urna, en cuyo caso no variará de forma apreciable la probabilidad de éxito entre una extracción y la siguiente. En esta situación, la distribución hipergeométrica se comportará aproximadamente como una distribución binomial; este hecho queda formalizado en el siguiente resultado.

Proposición 9.11 *Sea X una variable aleatoria con distribución hipergeométrica $\mathcal{H}(N,n,p)$. Cuando $N\to\infty$, $N_1\to\infty$ y $\frac{N_1}{N}\to p$, su función de probabilidad tiende hacia la de una binomial, $\mathcal{B}(n,p)$, es decir,*

$$\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}} p[X=k] = p[Y=k],$$

donde Y es una variable aleatoria con distribución binomial $\mathcal{B}(n,p)$.

Demostración.

$$p[X=k]=\frac{\binom{N_1}{k}\cdot\binom{N-N_1}{n-k}}{\binom{N}{n}},$$

probabilidad que, sin más que desarrollar los números combinatorios, es igual a

$$\binom{n}{k}\frac{N_1}{N}\cdot\frac{N_1-1}{N-1}\cdots\frac{N_1-k+1}{N-k+1}\cdot\frac{N-N_1}{N-k}\frac{N-N_1-1}{N-k-1}\cdots\frac{N-N_1-n+k+1}{N-n+1}.$$

Tomando límites,

$$\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}} p[X=k],$$

obtenemos:

$$\binom{n}{k}\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{N_1}{N}\cdot\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{\frac{N_1}{N}-\frac{1}{N}}{1-\frac{1}{N}}\cdots\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{\frac{N_1}{N}-\frac{k-1}{N}}{1-\frac{k-1}{N}}\cdot$$

$$\cdot\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{1-\frac{N_1}{N}}{1-\frac{k}{N}}\cdot\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{1-\frac{N_1}{N}-\frac{1}{N}}{1-\frac{k+1}{N}}\cdots\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}}\frac{1-\frac{N_1}{N}-\frac{n-k-1}{N}}{1-\frac{n-1}{N}},$$

es decir,

$$\lim_{\substack{N\to\infty\ N_1\to\infty\\ \frac{N_1}{N}\to p}} p[X=k]=\binom{n}{k}p^k(1-p)^{n-k}=p[Y=k]\ .\ \clubsuit$$

Más adelante, en la proposición 13.4, veremos cómo se puede extender esta convergencia a funciones de distribución.

En la práctica, la aproximación de una hipergeométrica a una binomial se considera buena cuando $N>50$ y $\frac{n}{N}\leq 0.1$.

Aunque existen tablas que proporcionan los valores de la función de distribución de una variable hipergeométrica, son poco utilizadas, por lo que no las presentamos en este libro[9].

9.8 Distribuciones discretas n-variantes

En esta sección estudiamos dos distribuciones discretas multivariantes. La multinomial, que es una generalización de la binomial cuando en un experimento interesa no la ocurrencia de un único suceso o la de su contrario, sino la de varios (tres o más). La segunda distribución que trataremos se denomina hipergeométrica multivariante y, como su nombre indica, es una generalización de la hipergeométrica, cuando las bolas de la urna en que se realizan las extracciones son de más de dos tipos.

9. El lector interesado puede consultarlas, por ejemplo, en *New Cambridge Elementary Statistical Tables*, Cambridge University Press, Cambridge, 1984.

9.8.1 Distribución multinomial

Sea Ω el espacio muestral asociado a un cierto experimento aleatorio, y supongamos que $\{A_i\}_{i=1}^n$ es una partición de Ω formada por n sucesos, es decir,

$$\Omega = A_1 \cup \cdots \cup A_n, \text{ tal que } A_i \cap A_j = \emptyset, \quad \text{para } i \neq j \ .$$

Obviamente, si $p_i = p(A_i)$, se cumple que $\sum_{i=1}^n p_i = 1$.

Se realizan m repeticiones independientes de dicho experimento. Consideremos n variables aleatorias $X_1, \ldots, X_n$, donde X_i es el *número de veces que se presenta el suceso A_i en los m experimentos* $(i = 1, \ldots, n)$. La variable n-dimensional $(X_1, \ldots, X_n)$ se dice que sigue una **distribución multinomial** de parámetros $m, p_1, \ldots, p_n$, que denotaremos por

$$(X_1, \ldots, X_n) \hookrightarrow \mathcal{M}(m, p_1, \ldots, p_n) \ .$$

Hallemos la probabilidad de obtener k_i veces cada resultado A_i, para $i = 1, \ldots, n$, siendo $k_1 + \cdots + k_n = m$. Uno de los sucesos para los que eso ocurre es aquel en que se obtiene el suceso A_1 en las k_1 primeras tiradas, A_2 en las k_2 siguientes y, finalmente, A_n en las últimas k_n tiradas. Esto es,

$$A_1 \ldots \overset{k_1}{\ldots} \ldots A_1 \ldots\ldots A_n \ldots \overset{k_n}{\ldots} \ldots A_n \ .$$

Dado que las repeticiones del experimento son independientes, la probabilidad del suceso anterior es

$$p_1^{k_1} \ldots p_n^{k_n} \ .$$

El número de sucesos en los que cada resultado A_i aparece k_i veces, para $i = 1, \ldots, n$, es

$$PR_m^{k_1, \ldots, k_n} = \frac{m!}{k_1! \cdots k_n!} \ ;$$

en consecuencia, y teniendo en cuenta que todos los sucesos de este tipo son equiprobables, la **función de probabilidad** de una variable aleatoria multinomial resulta:

$$f_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) = p[X_1 = k_1, \ldots, X_n = k_n] = \tfrac{m!}{k_1! \cdots k_n!} \cdot p_1^{k_1} \cdots p_n^{k_n} \ ,$$

para $k_i = 0, 1, \ldots, m$, con $k_1 + \cdots + k_n = m$.

La anterior función es, en efecto, una función de probabilidad, ya que toma valores positivos y, además[10],

$$\begin{aligned} \sum_{k_1} \cdots \sum_{k_n} p[X_1 = k_1, \ldots, X_n = k_n] &= \sum_{k_1} \cdots \sum_{k_n} \frac{m!}{k_1! \cdots k_n!} \cdot p_1^{k_1} \cdots p_n^{k_n} = \\ &= (p_1 + \cdots + p_n)^m = 1^m = 1 \ , \end{aligned}$$

donde $k_1 + \cdots + k_n = m$, para todos los sumandos.

10. Téngase en cuenta el desarrollo de Newton de la potencia entera de un polinomio:

$$(a_1 + \cdots + a_n)^m = \textstyle\sum_{k_1} \cdots \sum_{k_n} \frac{m!}{k_1! \cdots k_n!} \cdot a_1^{k_1} \cdots a_n^{k_n} \ ,$$

con $m = k_1 + \cdots + k_n$.

Un ejemplo de variable multinomial es el número de bolas blancas, negras y rojas obtenido al extraer diez bolas con reposición de una urna que contenga bolas de dichos colores en unas ciertas proporciones.

Ejemplo 9.8. El entrenador del equipo opina que sus jugadores A, B y C tienen aptitudes muy similares para formar parte del quinteto titular en la posición de base. Por ello, ha determinado que jueguen el mismo número de minutos en cada encuentro. Se sabe que el 30 % de las canastas de dos puntos, logradas entre los tres en un partido, corresponde al jugador A, mientras que B y C consiguen el 30 % y el 40 %, respectivamente.

Calculemos la probabilidad de que en un partido con un total de 9 encestes de dos puntos para los tres jugadores, A haya conseguido dos, B tres y C cuatro.

Consideremos la variable tridimensional (X_1, X_2, X_3), donde cada una de sus componentes representa el número de encestes de los jugadores A, B y C, respectivamente. La variable (X_1, X_2, X_3) sigue una distribución multinomial con $n = 9$, $p_1 = 0.3$, $p_2 = 0.3$ y $p_3 = 0.4$. Por tanto, la referida probabilidad es:

$$p[X_1 = 2, X_2 = 3, X_3 = 4] = \frac{9!}{2!3!4!} \cdot 0.3^2 \cdot 0.3^3 \cdot 0.4^4 = 0.078 \; . \triangleleft$$

Si $(X_1, \ldots, X_n)$ es una variable con distribución multinomial $\mathcal{M}(m, p_1, \ldots, p_n)$, existe una relación lineal entre sus componentes; en concreto, $X_1 + \cdots + X_n = m$. Por tanto, una de las variables, por ejemplo X_n, se puede poner como combinación lineal del resto, $X_n = m - X_1 - \cdots - X_{n-1}$. Entonces, el fenómeno que describe la variable $(X_1, \ldots, X_n)$ queda igualmente descrito por una variable de una dimensión menor, $(X_1, \ldots, X_{n-1})$, sin que esta pérdida de dimensión suponga una pérdida de información.

Una variable multinomial de dimensión dos, $\mathcal{M}(m, p_1, p_2)$, se puede describir, entonces, considerando una cualquiera de sus componentes. Ahora bien, cada componente, por ejemplo X_1, recoge el *número de veces que ocurre* A_1 en los m experimentos de Bernoulli independientes, es decir, X_1 es una binomial $\mathcal{B}(m, p_1)$. Con todo, la variable aleatoria binomial es unidimensional, y no bidimensional. Lo que ocurre es que, si (X_1, X_2) es una variable multinomial, quienes son binomiales son X_1 y X_2; pero, como hemos comentado anteriormente, X_1 y X_2 no se estudian conjuntamente, ya que $X_1 + X_2 = m$ y, en consecuencia, una de ellas se deduce a partir de otra.

Cada una de las n variables, X_i, que forman una multinomial, indica el número de veces que en las m tiradas aparece el suceso A_i; si sólo interesara esa variable, podríamos llamar éxito al suceso A_i, y fracaso a su contrario, $\overline{A}_i$, es decir, al suceso que representa el que ocurra uno cualquiera del resto de sucesos de la partición. El número de veces que se presenta dicho suceso, X_i, seguirá, obviamente, una distribución binomial. En la siguiente proposición se demuestra analíticamente la anterior deducción.

Proposición 9.12 *Sea* $(X_1, \ldots, X_n)$ *una variable aleatoria con distribución multinomial* $\mathcal{M}(m, p_1, \ldots, p_n)$*; entonces, la distribución marginal de* X_i *es binomial* $\mathcal{B}(m, p_i)$*, para cualquier* $i = 1, \ldots, n$.

Demostración. Hallemos la marginal de, por ejemplo, X_1:

$$\begin{aligned} p[X_1 = k_1] &= \sum_{k_2} \cdots \sum_{k_n} p[X_1 = k_1, \ldots, X_n = k_n] = \\ &= \frac{m!}{k_1!} \cdot p_1^{k_1} \sum_{k_2} \cdots \sum_{k_n} \frac{1}{k_2! \cdots k_n!} \cdot p_2^{k_2} \cdots p_n^{k_n} \end{aligned}$$

Como $k_2 + \cdots + k_n = m - k_1$ y $p_2 + \cdots + p_n = 1 - p$, tendremos que:

$$\begin{aligned} p[X_1 = k_1] &= \frac{m!}{k_1!(m-k_1)!} \sum_{k_2} \cdots \sum_{k_n} \frac{(m-k_1)!}{k_2! \cdots k_n!} \cdot p_2^{k_2} \cdots p_n^{k_n} = \\ &= \binom{m}{k_1} p_1^{k_1} (p_2 + \cdots + p_n)^{m-k_1} = \binom{m}{k_1} p_1^{k_1} (1-p_1)^{m-k_1} , \end{aligned}$$

que coincide con la función de probabilidad de una distribución binomial $\mathcal{B}(m, p_1)$. ♣

La generalización de esta proposición es inmediata, pudiéndose afirmar que cualquier grupo de variables de $(X_1, \ldots, X_n)$ sigue una distribución multinomial.

La siguiente proposición, que no demostraremos, proporciona un resultado análogo para las distribuciones condicionadas.

Proposición 9.13 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria con distribución multinomial. La distribución de cualquier grupo de variables de $(X_1, \ldots, X_n)$ condicionada por otro grupo de dichas variables, entre las que no se encuentren las primeras, también es multinomial.*

Calculemos la **función característica** de la distribución multinomial. Si aceptamos que la fórmula de Newton para el desarrollo de la potencia entera positiva de un polinomio real se verifica también en el campo complejo,

$$\begin{aligned} \varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) &= E\left[e^{i(t_1X_1 + \cdots + t_nX_n)}\right] = \\ &= \sum_{k_1} \cdots \sum_{k_n} e^{i(t_1k_1 + \cdots + t_nk_n)} \cdot \frac{m!}{k_1! \cdots k_n!} \cdot p_1^{k_1} \cdots p_n^{k_n} = \\ &= \sum_{k_1} \cdots \sum_{k_n} \frac{m!}{k_1! \cdots k_n!} \cdot (e^{it_1} p_1)^{k_1} \cdots (e^{it_n} p_n)^{k_n} = \\ &= (e^{it_1} p_1 + \cdots + e^{it_n} p_n)^m . \end{aligned}$$

Hallemos, por último, la esperanza y la matriz de varianzas-covarianzas de una variable aleatoria n-dimensional, $(X_1, \ldots, X_n)$, con distribución multinomial.

Dado que, como hemos visto, la distribución marginal de X_i es una binomial $\mathcal{B}(m, p_i)$, su esperanza es

$$E(X_i) = mp_i ,$$

y su varianza

$$Var(X_i) = mp_i(1 - p_i) .$$

Por tanto, la **esperanza** de $(X_1, \ldots, X_n)$ es la matriz

$$\begin{pmatrix} mp_1 \\ \vdots \\ mp_n \end{pmatrix} .$$

Para calcular la matriz de varianzas-covarianzas de $(X_1, \ldots, X_n)$ necesitaremos conocer la covarianza entre X_j y X_l. Obtengamos, previamente, $E(X_j \cdot X_l)$, derivando convenientemente la función característica. Así,

$$\frac{\partial^2}{\partial t_j \partial t_l} \varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) = m(m-1)\left(e^{it_1}p_1 + \cdots + e^{it_n}p_n\right)^{m-2} \cdot ie^{it_j}p_j \cdot ie^{it_l}p_l ,$$

de donde

$$\frac{\partial^2}{\partial t_j \partial t_l} \varphi_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n)\Big|_{(t_1,\ldots,t_n)=(0,\ldots,0)} = m(m-1)i^2 p_j p_l .$$

Por consiguiente,

$$E(X_j \cdot X_l) = m(m-1)p_j p_l ,$$

con lo cual

$$Cov(X_j, X_l) = m(m-1)p_j p_l - mp_j \cdot mp_l = -mp_j p_l .$$

Deducimos, de este modo, que la **matriz de varianzas-covarianzas** de la variable aleatoria n-dimensional, $(X_1, \ldots, X_n)$, es

$$\begin{pmatrix} mp_1(1-p_1) & -mp_1p_2 & \cdots & -mp_1p_n \\ -mp_2p_1 & mp_2(1-p_2) & \cdots & -mp_2p_n \\ \vdots & \vdots & \vdots & \vdots \\ -mp_np_1 & -mp_np_2 & \cdots & mp_n(1-p_n) \end{pmatrix} .$$

9.8.2 Distribución hipergeométrica multivariante

Consideremos una urna con N bolas ($N = 1, 2, \ldots$) de n colores, de las cuales N_1 son del primer color, N_2 son del segundo color y, finalmente, N_n son del n-ésimo color ($N = N_1 + \cdots + N_n$). Realizamos m extracciones sin reposición ($m \leq N$), y consideramos la variable, X_i, *número de bolas del color i extraídas* ($i = 1, \ldots, n$). La variable n-dimensional $(X_1, \ldots, X_n)$ se dice que sigue una **distribución hipergeométrica multivariante** de parámetros $N, m, p_1, \ldots, p_n$, donde $p_i = \frac{N_i}{N}$, es decir, la proporción de bolas del color i-ésimo en la primera extracción, para ($i = 1, 2, \ldots, n$). Esta distribución se designa abreviadamente como

$$(X_1, \ldots, X_n) \hookrightarrow \mathcal{H}(N, m, p_1, \ldots, p_n) .$$

Con un razonamiento similar al empleado en la obtención de la distribución hipergeométrica, se deduce su **función de probabilidad**,

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = p[X_1=k_1,\ldots,X_n=k_n] = \frac{\binom{N_1}{k_1}\cdots\binom{N_n}{k_n}}{\binom{N}{m}} = \\ = \frac{\binom{N\cdot p_1}{k_1}\cdots\binom{N\cdot p_n}{k_n}}{\binom{N}{m}},$$

donde $k_i = 0,1,\ldots,m$, con $k_1+\cdots+k_n=m$.

Comprobemos que la función de probabilidad está bien definida[11]. Por un lado, sus valores son no negativos; por otro,

$$\sum_{k_1}\cdots\sum_{k_n}\frac{\binom{N_1}{k_1}\cdots\binom{N_n}{k_n}}{\binom{N}{m}} = \frac{1}{\binom{N}{m}}\cdot\binom{N_1+\cdots+N_n}{k_1+\cdots+k_n} = \frac{\binom{N}{m}}{\binom{N}{m}} = 1\,.$$

Un ejemplo de esta variable es el número de *oros, copas, espadas* y *bastos* que recibe un jugador en una mano de mus.

Ejemplo 9.9. Los jugadores del equipo están molestos por las decisiones que ha tomado el entrenador en los últimos partidos, por lo cual han considerado oportuno comunicarle sus quejas mediante una comisión de 4 representantes elegidos al azar. Sabiendo que la plantilla del equipo está formada por 12 jugadores de los cuales 3 son pivotes, 3 son bases y 6 son aleros, ¿cuál es la probabilidad de que en esta comisión haya 2 bases y 2 pivotes?

Si consideramos la variable (X_1,X_2,X_3), donde cada componente recoge el número de pivotes, bases y aleros, respectivamente, que forman parte de la comisión, dicha variable tiene una distribución hipergeométrica multivariante con $N=12$, $n=4$, $N_1=3$, $N_2=3$ y $N_3=6$. Por tanto,

$$p[X_1=2,X_2=2,X_3=0] = \frac{\binom{3}{2}\binom{3}{2}\binom{6}{0}}{\binom{12}{4}} = 0.018\,. \triangleleft$$

Al igual que en la multinomial, las n variables que forman la hipergeométrica multivariante, $(X_1,\ldots,X_n)$, cumplen que $X_1+\cdots+X_n=m$. Por tanto, considerando una de ellas como combinación lineal del resto, por ejemplo $X_n = m-X_1-\cdots-X_{n-1}$, la podríamos eliminar de la expresión sin que ello acarrease una pérdida de información. Es decir, $(X_1,\ldots,X_n)$ y $(X_1,\ldots,X_{n-1})$ describen igualmente el experimento consistente en extraer sin devolución m bolas de n colores distintos. Asimismo, una hipergeométrica multivariante bidimensional queda perfectamente descrita por una cualquiera de sus componentes que es hipergeométrica

11. Será necesario recordar que, siendo $m=k_1+\cdots+k_n$

$$\sum_{k_1}\cdots\sum_{k_n}\binom{N_1}{k_1}\cdots\binom{N_n}{k_n} = \binom{N_1+\cdots+N_n}{m}\,.$$

unidimensional, pudiendo hacerse las mismas consideraciones que hicimos para la distribución multinomial.

Si sólo estuviéramos interesados en las bolas de un color, podríamos considerar que las demás son iguales entre sí (puede imaginar el lector que pinta éstas del mismo color), por lo que el número de aquéllas extraído seguirá una distribución hipergeométrica. La siguiente proposición confirma este resultado.

Proposición 9.14 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria con distribución hipergeométrica multivariante $\mathcal{H}(N, m, p_1, \ldots, p_n)$; entonces, la distribución marginal de X_i es hipergeométrica univariante, es decir, para cualquier $i = 1, \ldots, n$,*

$$X_i \hookrightarrow \mathcal{H}\left(N, m, p_i = \frac{N_i}{N}\right) .$$

Demostración. Hallemos la marginal de X_1, teniendo en cuenta que $k_2 + \cdots + k_n = m - k_1$ y $N_2 + \cdots + N_n = N - N_1$,

$$\begin{aligned} p[X_1 = k_1] &= \sum_{k_2} \cdots \sum_{k_n} p[X_1 = k_1, \cdots, X_n = k_n] = \\ &= \frac{\binom{N_1}{k_1}}{\binom{N}{m}} \cdot \sum_{k_2} \cdots \sum_{k_n} \binom{N_2}{k_2} \cdots \binom{N_n}{k_n} = \frac{\binom{N_1}{k_1}}{\binom{N}{m}} \cdot \binom{N_2 + \cdots + N_n}{k_2 + \cdots + k_n} = \\ &= \frac{\binom{N_1}{k_1} \cdot \binom{N - N_1}{m - k_1}}{\binom{N}{m}} , \end{aligned}$$

función de probabilidad de una distribución hipergeométrica de parámetros N, m y $p_1 = \frac{N_1}{N}$. ♣

Puede ampliarse la proposición anterior en el sentido de que cualquier grupo de variables extraído de $(X_1, \ldots, X_n)$ sigue una distribución hipergeométrica multivariante. La siguiente proposición, que no demostraremos, enuncia un resultado análogo para las distribuciones condicionadas.

Proposición 9.15 *Sea $(X_1, \ldots, X_n)$ una variable con distribución hipergeométrica multivariante. La distribución de cualquier grupo de variables de $(X_1, \ldots, X_n)$ condicionada por otro grupo de dichas variables, entre las que no se encuentren las primeras, también es hipergeométrica multivariante.*

Demos un resultado asintótico, que relaciona las distribuciones multinomial e hipergeométrica multivariante.

Proposición 9.16 *Sea $(X_1, \ldots, X_n)$ una variable aleatoria con distribución hipergeométrica multivariante. Cuando $N \to \infty$, $N_i \to \infty$ y $\frac{N_i}{N} \to p$, su función de probabilidad tiende a la de una multinomial $\mathcal{M}(m, p_1, \ldots, p_n)$.*

Supongamos una urna con bolas de n colores. Realizamos extracciones de esta urna y consideramos la variable n-dimensional $(X_1, \ldots, X_n)$, donde X_i designa

el *número de bolas elegidas del color i-ésimo.* Esta variable tiene una distribución multinomial si las extracciones son con reemplazamiento, y una distribución hipergeométrica multivariante si las extracciones son sin reemplazamiento. La proposición 9.16 muestra un resultado análogo al que vimos en la proposición 9.11 para las distribuciones univariantes binomial e hipergeométrica; afirma que, si la cantidad de bolas de cada color es muy grande y mucho mayor que el número de extracciones, el hecho de que la bola extraída no sea devuelta a la urna no modifica esencialmente las proporciones de bolas de cada color, es decir, las probabilidades que se obtienen, en este caso, son muy parecidas a las que tendríamos si la extracción se hiciera con devolución. En esta situación, la distribución hipergeométrica multivariante se comporta aproximadamente igual que una distribución multinomial.

Por último, dado que la marginal de una distribución hipergeométrica multivariante es una hipergeométrica, la esperanza de X_i es

$$E(X_i) = mp_i$$

y la varianza

$$Var(X_i) = mp_i(1-p_i)\frac{N-m}{N-1}\,.$$

Además, aunque no lo demostraremos, la covarianza entre X_i y X_j es

$$Cov(X_i, X_j) = -m\frac{N-m}{N-1}\sqrt{p_ip_j\frac{(1-p_i)(1-p_j)}{(N-N_i)(N-N_j)}N_iN_j}\,.$$

Así, la **esperanza** de la variable aleatoria n-dimensional $(X_1, \ldots, X_n)$ con distribución hipergeométrica multivariante es

$$\begin{pmatrix} mp_1 \\ \vdots \\ mp_n \end{pmatrix}$$

y su matriz de **varianzas-covarianzas** es

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix},$$

donde

$$\sigma_{ij} = \begin{cases} mp_i(1-p_i)\frac{N-m}{N-1} & \text{si } i = j \\ -m\frac{N-m}{N-1}\sqrt{p_ip_j\frac{(1-p_i)(1-p_j)}{(N-N_i)(N-N_j)}N_iN_j} & \text{si } i \neq j \end{cases}$$

9.9 Ejercicios propuestos

9.1. Se toma un número al azar, X, entre los cuatro primeros números naturales, 1, 2, 3 y 4. Elegido x, se hace una segunda elección, Y, entre los números naturales menores o iguales que x.

a) Sabiendo que la segunda elección ha sido el número 2, hállese el valor esperado en la primera elección.

b) Obténgase el coeficiente de correlación entre X e Y.

9.2. Dos amigos han adquirido sendos coches teledirigidos de la misma marca y modelo. El departamento de control de calidad de la empresa fabricante ha estimado que estos coches tienen una probabilidad de fallo de 0.05. ¿Cuál es la probabilidad de que terminen sin incidencias en el vehículo, una carrera a la que ambos se han apuntado?

9.3. Un teleadicto ha enviado una postal a un programa de una cierta cadena de televisión. Una azafata elige al azar la tarjeta agraciada con un premio de un millón de pesetas, entre quinientas mil postales. Suponiendo que cada postal, incluyendo el sello, ha costado 100 pesetas, ¿cuál es la ganancia esperada del referido concursante?

9.4. Si X es una variable aleatoria con distribución de Bernoulli de parámetro p, ¿para qué valor de p se maximiza $Var(X)$?

9.5. Sea X una variable con distribución binomial de parámetros $p = 0.3$ y $n = 15$. Hállese:

a) El mínimo valor de x tal que $p[X \leq x] \geq 0.3$.

b) El máximo valor de x tal que $p[X \geq x] \geq 0.4$.

9.6. ¿Cuál es el mínimo número de bolas rojas y blancas que han de colocarse en una bolsa para que la probabilidad de no sacar blanca coincida con la de obtener exactamente una bola blanca en 3 extracciones con reemplazamiento?

9.7. En un examen tipo test de 20 preguntas, cada una de ellas tiene 4 respuestas posibles. El alumno debe marcar una y sólo una de ellas. Cada respuesta acertada vale 1 punto y cada respuesta errónea resta 0.25 puntos. Es obligatorio contestar a todas las preguntas. El examen resulta aprobado si la puntuación final es de al menos 10 puntos.

Diez alumnos poco aventajados se ponen de acuerdo para responder al azar a todas las preguntas. ¿Cuál es la probabilidad de que al menos uno de ellos apruebe? Calcúlese esta probabilidad para veinte alumnos. A la vista de los resultados, ¿les interesa a los diez primeros alumnos buscar otros diez colaboradores?

9.8. Sean X_1 y X_2 variables aleatorias independientes binomiales $\mathcal{B}(n_1, p)$ y $\mathcal{B}(n_2, p)$. Hállese la distribución de X_1 condicionada a que $X_1 + X_2 = n$.

9.9. Sean X e Y variables aleatorias tales que X sigue una distribución $\mathcal{B}(m,p)$ y

$$p[Y=y \mid X=x] = \binom{n}{y-x} p^{y-x}(1-p)^{n-y+x} \qquad y = x, x+1, \ldots, x+n .$$

a) Hállese la distribución de la variable Y.

b) Calcúlese la distribución de X condicionada a $Y=y$.

9.10. Sea X una variable aletoria de Poisson de parámetro λ. Estúdiese el crecimiento y decrecimiento de su función de probabilidad y calcúlese su moda.

9.11. Un dispositivo está compuesto por un gran número de elementos que trabajan independientemente y con igual probabilidad de fallo durante el tiempo t.

a) Hállese la media y la desviación típica del número de elementos que fallan en el tiempo t, si la probabilidad de que en ese tiempo falle por lo menos un elemento es 0.98.

b) La probabilidad de que fallen a lo sumo dos elementos en el tiempo t.

9.12. El número de accidentes de trabajo, X, que se producen en una fábrica por semana sigue una distribución de Poisson. Sabiendo que el porcentaje de semanas en que ocurre un accidente es la mitad del correspondiente a las semanas en que no sucede ninguno, calcúlese:

a) El número esperado de accidentes semanales.

b) La probabilidad de que en una semana haya dos accidentes y en la siguiente otros dos.

c) La probabilidad de que en 4 semanas haya a lo sumo 8 accidentes.

d) La Dirección General de Trabajo decide declarar *semanas laborales blancas* aquellas en las que a lo sumo se produce un accidente. Si se considera un período de 10 semanas, determínese la probabilidad de que, como mínimo, resulten 2 *semanas blancas*.

9.13. Si X es una variable de Poisson de parámetro λ y $p[X=0] = \frac{1}{2}$, calcúlese $Var(X)$.

9.14. El número de clientes que llega diariamente a un establecimiento sigue una distribución de Poisson. Sabiendo que en una semana (6 días) acudieron 72 personas, hállese el número de clientes esperados el primer día de la semana.

9.15. El número de nuevas pólizas de un cierto tipo que una compañía de seguros suscribe en un año es una variable de Poisson de media 0. La probabilidad de que una póliza sea renovada para el siguiente año es 0.6. ¿Cuántas pólizas espera renovar la compañía?

9.16. Se observa una variable aleatoria, X, con distribución de Poisson de parámetro λ. Si el resultado es x, se lanza una moneda x veces, siendo p la probabilidad de

cara en dicha moneda. Sea Y la variable aleatoria que expresa el número de caras observadas.

a) Hállese la distribución de la variable Y, su media y su varianza.

b) Supongamos que hemos observado y caras, ¿cuál es la probabilidad de que hayamos tirado la moneda x veces?

9.17. Sea la variable aleatoria N con distribución de Poisson de parámetro λ. Sean $X_1, \ldots, X_N$, variables aleatorias independientes e igualmente distribuidas con distribución de Bernoulli de parámetro p e independientes de la variable N. Demuéstrese que $\sum_{i=1}^{N} X_i$ sigue una distribución de Poisson de parámetro λp.

9.18. En estudios demográficos sobre matrimonios que siguen algún tipo de planificación familiar se ha comprobado que el número, X, de hijos por matrimonio, sigue la siguiente distribución:

- En el 50 % de los casos $X = 1 - Y$, donde Y es una variable de Bernoulli de parámetro $p = 0.3$.
- En el resto de los casos $X = 2 + Z$, donde Z sigue una distribución de Poisson de parámetro λ.

a) Sabiendo que el número esperado de hijos por matrimonio es 2, hállese el valor de λ.

b) Se toma una muestra de 10 matrimonios. Calcúlese la probabilidad de que más de la mitad tenga al menos un hijo.

9.19. El número de pacientes atendidos cada día en el servicio de urgencias del hospital provincial es una distribución de Poisson, siendo 10 el número medio de llegadas. El porcentaje de pacientes que requieren hospitalización es del 40 %.

a) El hospital provincial cuenta un día determinado con 5 camas disponibles. Hállese la probabilidad de que el servicio de hospitalización se colapse.

b) Se ha calculado que el coste de hospitalización es de 10000 pesetas, mientras que el coste por paciente no hospitalizado sólo asciende a 5000. Hállese el coste esperado a consecuencia de las urgencias de un día determinado.

c) En la ciudad funciona independientemente una residencia sanitaria cuyo servicio de urgencias atiende diariamente a 12 pacientes por término medio, siendo la mitad de ellos hospitalizados. Un inspector sanitario comprueba que un día han sido hospitalizadas 16 personas entre ambos hospitales. Calcúlese la probabilidad de que al menos la mitad lo hayan sido en la residencia sanitaria.

9.20. Demuéstrese que, si X_1 y X_2 son variables aleatorias independientes con distribución geométrica de parámetro p, la distribución de X_1 condicionada por $X_1 + X_2 = n$ es una uniforme discreta de parámetro $n + 1$.

9.21. En una urna hay cinco fichas numeradas del 1 al 5. Las fichas se extraen al azar de una en una y con reemplazamiento. La selección continúa hasta que cada ficha ha sido extraída al menos una vez. Si han sido elegidas $i - 1$ de las 5 fichas,

llamaremos X_i al número de extracciones adicionales necesarias para sacar una ficha no seleccionada previamente ($i = 1, \ldots, 5$).

a) Demuéstrese que la variable $Y_i = X_i - 1$, $i = 1, \ldots, 5$, tiene una distribución geométrica.
b) Calcúlese $E(X_i)$, $i = 1, \ldots, 5$.
c) Consideremos la variable $W = X_1 + X_2 + \cdots + X_5$. Hállese $E(W)$, número medio de extracciones para sacar cada ficha al menos una vez.

9.22. En una fábrica se empaquetan cerillas en cajas mediante una máquina que no puede ser totalmente controlada. Para no perder clientes, la máquina se ajusta con el fin de que todas las cajas contengan al menos n cerillas. El número de cerillas en cada caja es una variable aleatoria discreta, X, cuya distribución viene dada por:

$$p[X = x] = (1 - p)p^{x-n} \qquad x \geq n \quad 0 < p < 1 \,.$$

El número de cerillas defectuosas en una caja que contiene m fósforos es una binomial de parámetros m y α, donde $0 < \alpha < 1$. Hállese el número medio de cerillas defectuosas.

9.23. Un experto en tiro con arco tiene una probabilidad de 0.9 de dar en la diana.

a) Sabiendo que en una serie de 10 tiros ha obtenido 7 aciertos, ¿cuál es la probabilidad de que el quinto fuera diana?
b) ¿Cuál es el número medio de tiros que debe realizar para completar una serie de 15 aciertos?

9.24. En el niquelado de ciertas láminas metálicas se producen desperfectos que se distribuyen aleatoriamente sobre toda la superficie niquelada. El número de defectos por lámina es una variable aleatoria de Poisson de media 2.

a) Se eligen al azar 10 láminas de la cadena de montaje. ¿Cuál es la probabilidad de que exactamente 3 de ellas tengan defectos?
b) Se toman láminas de la cadena de montaje hasta haber obtenido 7 sin defecto. ¿Cuál es la probabilidad de que haya habido que elegir diez láminas?
c) ¿Qué porcentaje de láminas niqueladas tendrá al menos 4 defectos?

9.25. Sea X una variable aleatoria con distribución hipergeométrica. Hállese su función de probabilidad, utilizando la regla de Laplace.

9.26. Se recibe un pedido de mil cajas con 100 tornillos cada una. Una caja es aceptada si, elegida una muestra aleatoria de 10 tornillos, resultan 2 defectuosos. Si la caja que va a ser examinada tiene 3 unidades defectuosas, ¿cuál es la probabilidad de que se acepte?

9.27. Una pecera contiene 3 peces exóticos rollizos y 7 desnutridos. El gato Félix saca hambriento 3 peces de la pecera; los 3 son rollizos y Félix se prepara a comérselos. En ese mismo momento aparece el amo del gato, un probabilista famoso, y le dice: «Félix, si puedes repetir al menos 2 veces lo que has hecho, de 3 veces que lo intentes, te dejaré que los comas. Si no lo logras, comerás tu acostumbrada ración de caracoles troceados.» ¿Cuál es la probabilidad de que Félix coma peces?

9.28. A Rosa le gustan los chicos morenos de octavo. En su colegio hay dos clases, A y B, a las que les corresponde el 55 % y el 45 %, respectivamente, de los chicos morenos de octavo. Sin embargo, para que sean del agrado de Rosa, no es suficiente con que sean morenos; además, han de tener los ojos verdes, el pelo rizado o ambas cosas a la vez.
La probabilidad de que un chico moreno de la clase A tenga los ojos verdes es 0.05 y de que su pelo sea rizado es 0.75, mientras que, en la clase B, estas probabilidades son 0.01 y 0.68.
Se eligen 20 chicos morenos del curso de octavo; ¿cuál es la probabilidad de que a Rosa no le guste ninguno de ellos?

9.29. Se ha lanzado un dado al azar 20 veces. Si se sabe que el resultado ha sido seis en 5 ocasiones, ¿cuál es la probabilidad de haber obtenido 8 unos, 3 doses y 4 cincos?

9.30. Una urna contiene 4 bolas blancas, 5 negras y 3 rojas. Sean X_1, X_2 y X_3 el número de bolas blancas, negras y rojas, respectivamente, obtenido en 3 extracciones con reemplazamiento.

a) Hállese la función de probabilidad de (X_1, X_2).
b) Obténgase la distribución de (X_1, X_2), sabiendo que se extrajo una bola roja.
c) Represéntese gráficamente la función de probabilidad de X_1.

9.31. En un viaje de estudios de tres días a Wall Street, el profesor acompañante pregunta la lección cada uno de los tres días a un alumno escogido al azar, que puede ser el mismo todos los días. El 50 % de los alumnos son de primer curso, el 30 % de segundo y el 20 % de tercero. Sean X_1, X_2 y X_3 el número de alumnos de primero, segundo y tercero, respectivamente, a quienes se les pregunta la lección durante el viaje.

a) Demuéstrese que la variable X_1 tiene una distribución binomial.
b) Calcúlense los coeficientes de correlación entre X_1 y X_2 y entre X_1 y X_2+X_3.

9.32. Obténgase la función de probabilidad de una distribución hipergeométrica trivariante.

9.33. El consejo de un departamento universitario está formado por dos catedráticos, seis titulares, cuatro asociados, un becario y un alumno. Con objeto de realizar una revisión de planes de estudio, se elige al azar una comisión de cinco miembros. ¿Cuál es la probabilidad de que esté compuesta por un catedrático, dos titulares y un alumno?

9.34. Resuélvase el ejercicio 9.30 cuando las extracciones se hacen sin reemplazamiento.

Capítulo 10

MODELOS DE DISTRIBUCIONES CONTINUAS UNIVARIANTES

En este capítulo y en el siguiente estudiamos las distribuciones univariantes continuas. Como dijimos en la introducción del capítulo anterior, se trata de distribuciones tipo; su función de densidad depende, por tanto, de los valores que asignemos a ciertos parámetros de forma que, una vez elegido el tipo de distribución que mejor describe un fenómeno aleatorio desde el punto de vista probabilístico, se buscan los valores de los parámetros que permitan que la distribución se ajuste lo mejor posible a la información de que se disponga.

Las variables aleatorias unidimensionales que se estudian en este capítulo son las más utilizadas en las aplicaciones. Entre ellas, está la variable aleatoria normal, que desde hace dos siglos y medio se ha convertido en la clave, tanto de la modelización de los fenómenos aleatorios de tipo continuo, como de la Inferencia Estadística. De hecho, en el capítulo undécimo se estudian tres variables aleatorias que surgen cuando se realizan estimaciones o se contrastan hipótesis referentes a fenómenos aleatorios que siguen una distribución normal.

10.1 Distribución uniforme continua

Una variable aleatoria X se dice que sigue una **distribución uniforme continua** en el intervalo real (a, b), si su función de densidad es constante en dicho intervalo y nula fuera de él, es decir,

$$f_X(x) = \begin{cases} k & a < x < b \\ 0 & \text{en el resto} \end{cases}$$

Calculemos el valor de k; para ello, integraremos la función de densidad en el campo de definición de la variable, esto es,

$$\int_a^b k \cdot dx = k(b-a) = 1\,,$$

con lo que $k = \frac{1}{b-a}$. Por tanto, la **función de densidad** de una variable uniforme continua de parámetros a y b $(a, b \in \mathbb{R}, a < b)$ es

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{en el resto} \end{cases}$$

Expresaremos abreviadamente que X sigue una distribución uniforme en el intervalo (a, b) como

$$X \hookrightarrow \mathcal{U}(a, b) .$$

La representación gráfica de esta función de densidad es la mostrada en la figura 10.1.

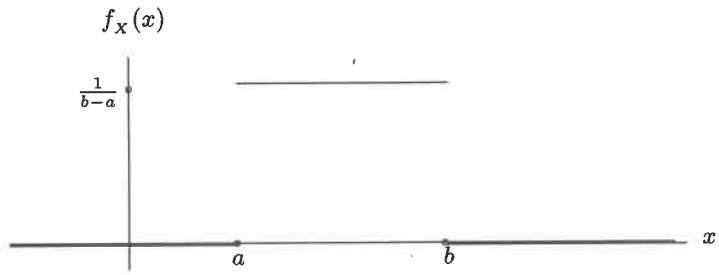


FIG. 10.1.

La distribución uniforme, también llamada rectangular por el aspecto de su función de densidad, se presenta de una forma tan natural que seguramente se viene empleando desde hace muchos años. Se tiene constancia de su utilización por T. BAYES en 1763 y por P. S. DE LAPLACE en 1812. Según la definición de esta variable, cualquier elección de números reales al azar y sin preferencias en un intervalo de longitud finita, es una variable aleatoria uniforme. Con frecuencia, esta distribución se asocia al comportamiento aleatorio de una magnitud de la que sólo se conoce su campo de variación (un intervalo finito).

Su **función de distribución** resulta ser

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

esto es, vale cero para $x \leq a$, la unidad para $x \geq b$, y entre a y b es un segmento rectilíneo (figura 10.2).

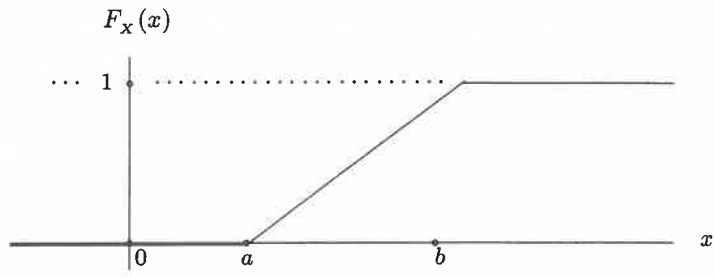


FIG. 10.2.

Ejemplo 10.1. Aunque Juan vive muy cerca de su lugar de trabajo, siempre va en coche. En el trayecto hay dos semáforos que se cierran cada 2 minutos, permaneciendo en esta situación durante medio minuto. Dichos semáforos están sincronizados de modo que el segundo de ellos se pone rojo a los 2.5 minutos de abrirse el primero.

Si Juan ha tenido que detenerse en el primero, y el tiempo en minutos que tarda en recorrer la distancia que los separa sigue una distribución uniforme en el intervalo $(1, 4)$, ¿cuál es la probabilidad de que tenga que pararse también en el segundo semáforo?

Para resolver esta cuestión basta considerar que, puesto que el segundo semáforo se cierra a los dos minutos y medio de ponerse verde el primero y permanece en rojo 0.5 minutos, Juan no tendrá más remedio que parar su coche cuando emplee un tiempo entre 2.5 y 3 minutos en ir de uno a otro. Es decir, denotando por X la variable que designa este tiempo, la probabilidad pedida es:

$$p[2.5 < X < 3] = F_X(3) - F_X(2.5) = \frac{3-1}{4-1} - \frac{2.5-1}{4-1} = 0.166 \,. \triangleleft$$

Los momentos de esta variable se calculan resolviendo integrales sencillas.

Esperanza:

$$E(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{b+a}{2} \,,$$

valor que será, también, el de la mediana.

Varianza:

Previamente calcularemos el momento de segundo orden,

$$E(X^2) = \int_a^b x^2 \cdot \frac{1}{b-a} dx = \frac{b^2+ab+a^2}{3} \,,$$

con lo cual la varianza es

$$Var(X) = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12} \,.$$

La **función característica** se obtuvo en el ejemplo 8.2 para un caso concreto, la distribución $\mathcal{U}(0,1)$. Un razonamiento análogo conduciría a la función característica de una distribución uniforme $\mathcal{U}(a,b)$.

$$\varphi_X(t) = E(e^{itX}) = \int_a^b e^{itx}\frac{1}{b-a}dx = \begin{cases} \frac{e^{itb}-e^{ita}}{it(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

Puesto que las variables aleatorias uniformes reparten la probabilidad por igual en todo el intervalo de variación, cabe esperar que si se realiza una transformación lineal de una variable uniforme (un cambio de origen y de escala, por ejemplo), la variable transformada siga siendo uniforme. La siguiente proposición confirma este resultado.

Proposición 10.1 *Sea X una variable con distribución uniforme en el intervalo (a,b). Entonces, si c y d son números reales con $c \neq 0$, la variable $Y = cX + d$ tiene una distribución uniforme en el intervalo $(ca+d, cb+d)$, si c es positivo, y en el $(cb+d, ca+d)$, si c es negativo.*

Si c es igual a cero, Y tiene una distribución degenerada en d.

Demostración. La última afirmación es evidente, pues, en este caso, toda la probabilidad de la variable se concentra en un punto, d. Demostremos la primera.

Cuando $y = cx + d$, la transformación inversa vale $x = \frac{y-d}{c}$. Por tanto,

$$\frac{dx}{dy} = \frac{1}{c}$$

y, aplicando la proposición 4.14, la función de densidad de Y resulta

$$f_Y(y) = \begin{cases} \frac{1}{b-a}|\frac{1}{c}| & a < \frac{y-d}{c} < b \\ \\ 0 & \text{en el resto} \end{cases}$$

Si $c > 0$, $a < \frac{y-d}{c} < b$ equivale a $ca + d < y < cb + d$, y si $c < 0$, equivale a $ca + d > y > cb + d$. Además, en el primer caso, $|c| = c$ y, en el segundo, $|c| = -c$. Sustituyendo en la expresión anterior, obtenemos que, si $c > 0$,

$$f_Y(y) = \begin{cases} \frac{1}{cb-ca} & ca + d < y < cb + d \\ \\ 0 & \text{en el resto} \end{cases}$$

y si $c < 0$,

$$f_Y(y) = \begin{cases} \frac{1}{ca-cb} & cb + d < y < ca + d \\ \\ 0 & \text{en el resto} \end{cases}$$

que son las funciones de densidad de las variables uniformes que indica el enunciado. ♣

Las tablas de números aleatorios (de las que presentamos una en el anexo final de este libro) recogen realizaciones independientes de una variable uniforme en el intervalo $(0, a)$, donde a es una cantidad que depende del número de cifras que tengan los números aleatorios. Así, si son de cinco cifras, como la que aparece en este libro, el valor de a será cien mil; si son de cuatro cifras, a valdrá diez mil.

Aplicando la proposición anterior, y a partir de las tablas de números aleatorios, se pueden obtener realizaciones de una variable uniforme cualquiera.

Los ordenadores y la mayor parte de las calculadoras de bolsillo proporcionan asimismo números aleatorios; pero, como realmente no los obtienen al azar, sino mediante una fórmula, la calidad de esos números deja, a veces, mucho que desear, en el sentido de que ciertas zonas del intervalo tienen mayores probabilidades que otras o, también, de que se observa un patrón de comportamiento que hace difícil aceptar la independencia de las realizaciones. Los fabricantes serios someten los procedimientos de generación de números aleatorios (y las tablas que a partir de ellos se obtienen) a diversos contrastes para medir su calidad.

Además, a partir de unas tablas de números aleatorios (de hecho, realizaciones independientes de una variable uniforme en el intervalo $(0, a)$), se pueden obtener números aleatorios que sean realizaciones de cualquier otra distribución de probabilidades. Para ello, se utiliza el siguiente resultado.

Proposición 10.2 *Sea X una variable aleatoria continua con función de distribución F_X estrictamente creciente en el campo de variación de X. La variable $Y = F_X(X)$ tiene una distribución uniforme en el intervalo $(0, 1)$.*

Demostración.

$$F_Y(y) = p[Y \leq y] = p[F_X(X) \leq y] \, .$$

Como F_X varía entre cero y uno por ser una función de distribución, esta probabilidad es cero para $y \leq 0$ y uno para $y \geq 1$. Por otro lado, si $0 < y < 1$,

$$F_Y(y) = p[F_X(X) \leq y] = p[X \leq F_X^{-1}(y)] = F_X(F_X^{-1}(y)) = y \, .$$

El hecho de que F_X sea continua y estrictamente creciente garantiza la existencia de F_X^{-1}. En definitiva,

$$F_Y(y) = \begin{cases} 0 & y \leq 0 \\ y & 0 < y < 1 \\ 1 & y \geq 1 \end{cases}$$

Su función de densidad la obtenemos derivando la anterior función de distribución

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{en el resto} \end{cases}$$

En consecuencia, $Y = F_X(X) \hookrightarrow \mathcal{U}(0, 1)$. ♣

El ejemplo siguiente ilustra cómo utilizar esta proposición para obtener realizaciones de una variable aleatoria continua a partir de una tabla de números aleatorios.

Ejemplo 10.2. Se sabe que el tiempo de vida, en años, de unas bombillas es una variable continua, X, con la función de densidad siguiente[1]:

$$f_X(x) = \begin{cases} 0.2e^{-0.2x} & x > 0 \\ 0 & \text{en el resto} \end{cases}$$

Deseamos simular el tiempo de vida de ocho bombillas, para mostrar a un comprador el comportamiento que de las mismas puede esperarse.

La función de distribución de X vale

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-0.2x} & x > 0 \end{cases}$$

Por tanto, aplicando la proposición anterior,

$$Y = F_X(X) = \begin{cases} 0 & X \leq 0 \\ 1 - e^{-0.2X} & X > 0 \end{cases}$$

es una variable uniforme en el intervalo $(0, 1)$. Obsérvese que, puesto que X es siempre positiva, podemos escribir esta transformación como

$$Y = 1 - e^{-0.2X} .$$

Consecuentemente,

$$X = F_X^{-1}(Y) ,$$

donde F_X^{-1} es la transformación inversa de la función de distribución de X. Sustituyendo F_X^{-1} por su expresión, esto es, despejando X, obtenemos

$$X = -5 \cdot \ln (1 - Y) .$$

Hemos escrito la variable aleatoria X, cuyo comportamiento queremos simular, como una función de una variable $\mathcal{U}(0, 1)$. Si disponemos de ocho realizaciones de una $\mathcal{U}(0, 1)$, podremos generar otras tantas de la variable X.

La tabla de números aleatorios que incluimos en el anexo final de este libro, proporciona realizaciones de una variable Z con distribución $U(0, 100000)$. Seleccionemos los ocho primeros.

$$\begin{array}{llll} z_1 = 49914 & z_2 = 68221 & z_3 = 56088 & z_4 = 06112 \\ z_5 = 95481 & z_6 = 30094 & z_7 = 68493 & z_8 = 88796 . \end{array}$$

Pero, aplicando la proposición 10.1, si $Z \hookrightarrow U(0, 100000)$, entonces $Y = \frac{Z}{100000}$ es $U(0, 1)$. Por tanto, si dividimos por cien mil los números aleatorios anteriores, tenemos ocho realizaciones de la variable Y,

$$\begin{array}{llll} y_1 = 0.49914 & y_2 = 0.68221 & y_3 = 0.56088 & y_4 = 0.06112 \\ y_5 = 0.95481 & y_6 = 0.30094 & y_7 = 0.68493 & y_8 = 0.88796 . \end{array}$$

1. En la sección 10.5 veremos que esta función de densidad corresponde a una distribución denominada exponencial negativa.

A partir de ellas, obtendremos las realizaciones de la variable X, que simulan la duración de las ocho bombillas. La primera durará:

$$x_1 = -5 \cdot \ln(1 - y_1) = -5 \cdot \ln(1 - 0.49914) = 1.50 \text{ años.}$$

La duración, en años, de las siete restantes será:

$$x_2 = 2.49 \quad x_3 = 1.79 \quad x_4 = 0.14 \quad x_5 = 6.73 \quad x_6 = 0.78 \quad x_7 = 2.51 \quad x_8 = 4.75 \; . \triangleleft$$

Si la variable X, cuyo comportamiento queremos simular, es discreta, el procedimiento anterior no es aplicable de la misma forma, ya que su función de distribución, F_X, es escalonada y, en consecuencia, no se puede despejar la variable X en función de Y. Con todo, la idea que subyace en la proposición 10.2 puede concretarse en un procedimiento similar.

Supongamos que X es una variable discreta, y que conocemos su función de probabilidad:

$$p[X = x_n] = p_n \; , \qquad \text{para } n = 1, 2, \ldots$$

Con lo cual, su función de distribución es

$$F_X(x) = \sum_{x_n \leq x} p[X = x_n] = \sum_{x_n \leq x} p_n \; ,$$

para todo número real x.

Obtengamos una realización, y, de una variable $U(0,1)$. Diremos que la realización correspondiente de la variable X es x_1 si

$$0 < y \leq F_X(x_1) \; ;$$

análogamente, si

$$F_X(x_{i-1}) < y \leq F_X(x_i) \; ,$$

diremos que la realización de X es el valor x_i $(i = 1, 2, 3, \ldots)$.

Se trata, por tanto, de dividir el intervalo $(0,1)$ en tantos intervalos como valores tome la variable X, de forma que si el número aleatorio uniforme en el intervalo $(0,1)$ se encuentra en $(0, p_1]$, decimos que se ha obtenido el valor x_1 de X; si se encuentra en $(p_1, p_1 + p_2]$, decimos que el valor obtenido es x_2; si se encuentra en $(p_1 + p_2, p_1 + p_2 + p_3]$, decimos que hemos obtenido el valor x_3, y así sucesivamente.

Ejemplo 10.3. Hallemos a partir de los ocho números aleatorios, $y_1, \ldots, y_8$ anteriores, la simulación de ocho lanzamientos independientes de un dado perfecto.

Sea X la variable que representa los resultados de un lanzamiento. Su función de probabilidad es

$$p[X = x] = \frac{1}{6} \; , \quad x = 1, 2, 3, 4, 5, 6 \; .$$

Entonces, si la variable aleatoria Y, uniforme en el intervalo $(0,1)$, se encuentra en el intervalo $(0, \frac{1}{6}]$, diremos que el resultado del lanzamiento, X, ha sido un uno. Si se encuentra en el $(\frac{1}{6}, \frac{2}{6}]$, diremos que el resultado ha sido un dos, y así sucesivamente. La tabla siguiente nos indica cómo asignar resultados al lanzamiento, según el número aleatorio obtenido.

Valor de Y	Valor de X
(0,0.1667]	1
(0.1667,0.3333]	2
(0.3333,0.5]	3
(0.5,0.6667]	4
(0.6667,0.8333]	5
(0.8333,1]	6

En concreto, para las ocho realizaciones de Y, $y_1, \ldots, y_8$, del ejemplo anterior, se obtiene la siguiente simulación de ocho lanzamientos independientes de un dado:

$$3, 5, 4, 1, 6, 2, 5, 6 \; . \triangleleft$$

10.2 Distribución normal

Una variable aleatoria se dice que sigue una **distribución normal** de parámetros μ y σ (con $\mu, \sigma \in \mathbb{R}$ y tales que $-\infty < \mu < \infty, \sigma > 0$), si su **función de densidad** es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad , \quad \text{para } -\infty < x < \infty \; .$$

Una variable aleatoria normal se designa por $N(\mu, \sigma)$ y también recibe el nombre de distribución de LAPLACE-GAUSS.

Veamos que la función de densidad está bien definida. Como, evidentemente, es mayor o igual que cero, nos bastará con comprobar que la integral en toda la recta real es la unidad; para ello, efectuaremos el cambio de variable $t = \frac{x-\mu}{\sqrt{2}\sigma}$, con lo cual

$$\int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{\pi}} \cdot \int_{t=-\infty}^{\infty} e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \cdot \int_{t=0}^{\infty} e^{-t^2} dt \; .$$

Si hacemos el cambio de variable $t^2 = u$ (y, por tanto, $2tdt = du$), esta integral vale

$$\int_{u=0}^{\infty} e^{-t^2} dt = \frac{1}{2} \int_{u=0}^{\infty} e^{-u} \cdot u^{-1/2} du = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \; ,$$

donde Γ es una de las funciones eulerianas[2]. Por tanto,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-1}{2\sigma^2}(x-\mu)^2} dx = \frac{2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = 1 \; .$$

2. La función euleriana gamma de p, $\Gamma(p)$, se define como

$$\Gamma(p) = \int_0^{\infty} x^{p-1} \cdot e^{-x} dx \qquad p > 0 \; .$$

Algunas de sus propiedades de interés son:

i) $\Gamma(1) = 1$.
ii) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.
iii) $\Gamma(p) = (p-1) \cdot \Gamma(p-1)$ con $p > 1$.
iv) $\Gamma(n) = (n-1)!$ con n entero.

Fue DE MOIVRE quien en 1733 estableció la expresión matemática de la distribución normal como aproximación de la distribución binomial. Posteriormente, LAPLACE y GAUSS la hallaron empíricamente, estudiando la distribución de los errores de medición y, tras sus trabajos, se convirtió en la distribución más utilizada en las distintas aplicaciones (de ahí su nombre). En 1774, LAPLACE demostró que también es una buena aproximación de la distribución hipergeométrica y, alrededor del año 1780, obtuvo unas tablas con los valores de su función de distribución. El calificativo *normal* se debe a que esta distribución es típica de muchos experimentos y observaciones, en especial de fenómenos de la naturaleza donde intervienen muchas causas, cada una de las cuales produce un pequeño efecto, y que pueden suponerse independientes.

Aunque a comienzos de este siglo la universalidad de su uso se puso en duda, su papel único en el estudio de las convergencias de sucesiones de variables aleatorias (teorema del límite central) se ha mantenido hasta nuestros días, con nuevos desarrollos que siguen proporcionando resultados de interés.

Estudiemos el comportamiento de la función de densidad de esta variable. Puede comprobar el lector que se cumplen las siguientes propiedades.

1) $f_X(x)$ es una función continua en toda la recta real.

2) $f_X(x)$ es simétrica respecto del parámetro μ, es decir,

$$f_X(\mu + t) = f_X(\mu - t) .$$

3) $f_X(x)$ nunca toma el valor cero, pero tiene una asíntota horizontal que es el eje de abscisas, esto es,

$$\lim_{x \to \infty} f_X(x) = 0$$

y

$$\lim_{x \to -\infty} f_X(x) = 0 .$$

4) $f_X(x)$ es estrictamente creciente en los puntos $x < \mu$ y es estrictamente decreciente en los puntos $x > \mu$. Para comprobarlo basta estudiar su derivada

$$f'_X(x) = \frac{-(x-\mu)}{\sigma^2} \cdot f_X(x) ,$$

que será positiva cuando $x < \mu$, y negativa si $x > \mu$.

5) $f_X(x)$ presenta un máximo absoluto en el punto μ, y el máximo vale $\frac{1}{\sqrt{2\pi}\sigma}$. En efecto, la primera derivada se anula para $x = \mu$. Calculando la segunda derivada,

$$f''_X(x) = \frac{-1}{\sigma^2} \cdot f_X(x) + \left(\frac{x-\mu}{\sigma^2}\right)^2 \cdot f_X(x) ,$$

la cual, en $x = \mu$, punto de anulación de la primera derivada, toma un valor negativo,

$$f''_X(\mu) = \frac{-1}{\sigma^2} \cdot f_X(\mu) ,$$

por lo que $f_X(x)$ alcanza un máximo relativo en $x = \mu$ y, además, $f_X(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$. Dicho máximo relativo es un máximo absoluto por la propiedad anterior.

6) $f_X(x)$ tiene dos puntos de inflexión en $x = \mu + \sigma$ y en $x = \mu - \sigma$, como se comprueba de inmediato, igualando a cero la segunda derivada y estudiando la tercera.

La figura 10.3 muestra la representación gráfica de dicha función.

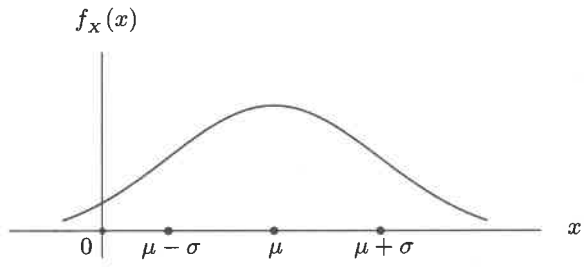


FIG. 10.3.

Obsérvese que el parámetro μ proporciona el centro de la distribución, mientras que σ indica el mayor o menor grado de *apertura* de la curva que, por su aspecto, se suele llamar *campana de Gauss*; por tanto, μ es un parámetro de centralización, y σ de dispersión o de concentración. Este hecho se confirmará, posteriormente, cuando demostremos que la esperanza y la desviación típica de la distribución normal son, precisamente, μ y σ. La figura 10.4 ilustra esto, representando conjuntamente la función de densidad de dos distribuciones normales: $X_1 \hookrightarrow N(\mu_1, \sigma_1)$ y $X_2 \hookrightarrow N(\mu_2, \sigma_2)$, donde $\mu_1 < \mu_2$ y $\sigma_1 > \sigma_2$.

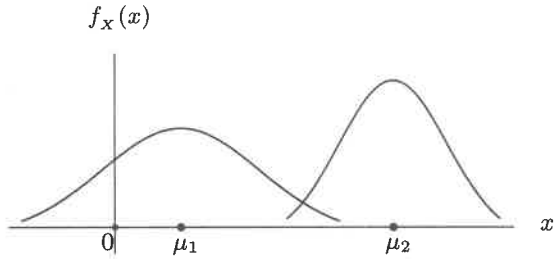


FIG. 10.4.

Una de las más importantes, dentro de la familia de distribuciones normales, es la distribución normal de parámetros $\mu = 0$ y $\sigma = 1$, es decir, la distribución

$N(0,1)$. Esta distribución, que se suele designar por Z, se conoce como **distribución normal estándar** o **tipificada**, y su función de densidad es

$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \qquad -\infty < z < \infty .$$

Su importancia deriva del hecho de que, como comprobaremos en la proposición siguiente, se puede obtener una variable aleatoria normal cualquiera a partir de una $N(0,1)$.

Proposición 10.3 *Sea X es una variable aleatoria con distribución $N(\mu,\sigma)$. Entonces, $\frac{X-\mu}{\sigma}$ es una variable con distribución $N(0,1)$.*

Demostración. Denotemos por Z a la variable $\frac{X-\mu}{\sigma}$, y calculemos su función de distribución,

$$F_Z(z) = p[Z \leq z] = p[X \leq \sigma z + \mu] = F_X(\sigma z + \mu) .$$

Derivándola, obtenemos que

$$f_Z(z) = f_X(\sigma z + \mu) \cdot \sigma = \frac{\sigma}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2\sigma^2}(\sigma z + \mu - \mu)^2} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} ,$$

función de densidad de una distribución $N(0,1)$. ♣

En la subsección 6.2.3 definimos la tipificación de una variable como el cambio de origen y escala que se realiza cuando a una variable aleatoria le restamos su esperanza y, posteriormente, la dividimos por su desviación típica. Como veremos más adelante, la esperanza de una variable $N(\mu,\sigma)$ es su primer parámetro, μ, y la desviación típica es el segundo, σ. La proposición anterior indica, por tanto, que el adjetivo *tipificada* que hemos utilizado para designar a la distribución $N(0,1)$ está bien empleado, pues esta variable resulta de tipificar una normal cualquiera. Evidentemente, puede realizarse el proceso contrario, recogido en la siguiente proposición, y cuya demostración dejamos a cargo del lector.

Proposición 10.4 *Sea Z una variable aleatoria con distribución $N(0,1)$. La variable $X = \sigma Z + \mu$, donde μ y σ son dos números reales cualesquiera, con $\sigma > 0$, sigue una distribución $N(\mu,\sigma)$.*

La **función de distribución** de la normal $N(\mu,\sigma)$ es

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-1}{2\sigma^2}(t-\mu)^2} dt \qquad -\infty < x < \infty .$$

No se conoce la primitiva de la función de densidad de la distribución normal, por lo que los valores de su función de distribución representada en la figura 10.5, se calculan por procedimientos numéricos aproximados.

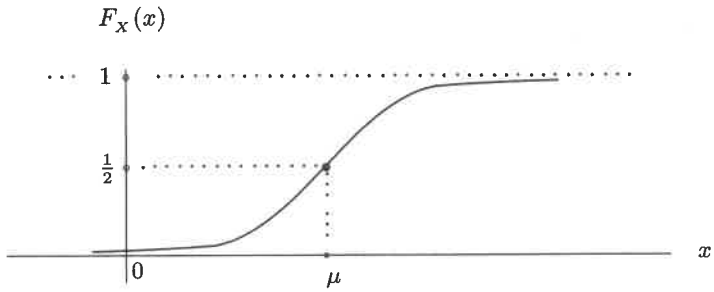


FIG. 10.5.

Como veremos en el ejemplo 10.4, la función de distribución de la normal estándar,

$$F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-t^2}{2}} dt \qquad -\infty < z < \infty \,,$$

está tabulada; esta función la denotaremos como $\Phi(z)$. A partir de ella, y utilizando la proposición 10.3, se pueden obtener probabilidades para una variable normal cualquiera.

La figura 10.6 representa el área bajo la curva de la función de densidad normal estándar en el intervalo $(-\infty, z)$; obviamente, el valor de esa área es la función de distribución de la $N(0,1)$ en z, es decir, $\Phi(z)$.

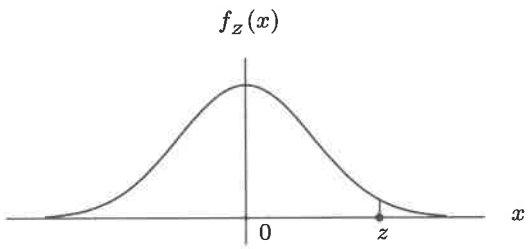


FIG. 10.6.

Aunque la tabla de la distribución normal tipificada no incorpora los valores de $\Phi(z)$ cuando z es negativo, ello no representa problemas dada la simetría de la función de densidad respecto del cero. Así, si $z < 0$,

$$\Phi(z) = p[Z \leq z] = p[Z \geq -z] = 1 - p[Z < -z] = 1 - \Phi(-z) \,,$$

donde $\Phi(-z)$ sí aparece en las tablas por ser $-z$ un valor positivo.

Los momentos de una variable normal se hallan, sin dificultad, a partir de la función característica, función que para esta distribución es una herramienta sencilla de manejar y muy útil. Obtengamos, pues, la función característica de una variable $N(\mu, \sigma)$, calculando, para ello, en primer lugar, la de una $N(0,1)$.

Función característica:
Sea $Z \hookrightarrow N(0,1)$; su función característica es

$$\begin{aligned}\varphi_z(t) = E(e^{itZ}) &= \int_{z=-\infty}^{\infty} e^{itz} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} dz = \\ &= e^{\frac{i^2t^2}{2}} \cdot \int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-(z^2-2itz+i^2t^2)}{2}} dz = \\ &= e^{\frac{-t^2}{2}} \cdot \int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-(z-it)^2}{2}} dz = \\ &= e^{\frac{-t^2}{2}} \cdot \int_{u=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-u^2}{2}} du = e^{\frac{-t^2}{2}} ,\end{aligned}$$

sin más que hacer el cambio de variable[3] $u = z - it$.

Para obtener la función característica de una variable aleatoria X con distribución $N(\mu, \sigma)$, basta considerar que, por la proposición 10.4, $X = \sigma \cdot Z + \mu$, donde Z es una normal estándar; por tanto,

$$\varphi_x(t) = E(e^{itX}) = E\left[e^{it(\sigma Z+\mu)}\right] = e^{it\mu} \cdot \varphi_z(t\sigma) = e^{it\mu - \frac{t^2\sigma^2}{2}} .$$

Calculemos, ahora, los dos primeros momentos.

Esperanza:
Derivando la función característica, tendremos que

$$\varphi'_x(t) = (i\mu - t\sigma^2) \cdot \varphi_x(t) ,$$

con lo cual, y dado que para cualquier variable aleatoria X, $\varphi_x(0) = 1$,

$$E(X) = \frac{\varphi'_x(0)}{i} = \mu .$$

La esperanza coincide, por tanto, con el primer parámetro de la distribución. Obsérvese, además, que en una distribución $N(\mu, \sigma)$, la moda, la mediana y la media o esperanza, coinciden. En efecto, por la propiedad 5, el máximo de la función de densidad se obtiene en el punto μ que es, consecuentemente, la moda de la distribución. Por otro lado, la propiedad 2 indica que la función de densidad es simétrica con respecto a μ; entonces,

$$p[X \leq \mu] = p[X \geq \mu] .$$

Pero ambas probabilidades suman la unidad, por lo que $p[X \leq \mu] = 0.5$. El parámetro μ es, en consecuencia, la mediana de la distribución.

3. Adviértase que $u = z - it$ es un número complejo y, como ya hemos comentado en el capítulo 8, la última integral no puede obtenerse como una integral real. Sin embargo, aunque los cálculos de funciones características que se realizan en este capítulo no son formales, ilustran al lector sobre el aspecto de dichas funciones.

Varianza:
Derivando otra vez la función característica,

$$\varphi_X''(t) = -\sigma^2 \cdot \varphi_X(t) + (i\mu - t\sigma^2)^2 \cdot \varphi_X(t) ,$$

con lo cual,

$$E(X^2) = \frac{\varphi_X''(0)}{i^2} = \sigma^2 + \mu^2 ,$$

y la varianza es

$$Var(X) = \sigma^2 .$$

Ejemplo 10.4. La empresa de Juan se dedica a la fabricación de maquinaria industrial. Entre la piezas que se elaboran están unos tubos de acero cuyos diámetros interiores tienen una distribución normal de media 5 centímetros y desviación típica 0.5. ¿Qué porcentaje de tubos tienen diámetro superior a 6 centímetros?
La probabilidad pedida es

$$p[X > 6] = 1 - p[X \leq 6] ,$$

donde X denota la *longitud, en centímetros, del diámetro interior de un tubo cualquiera.* Sin más que tipificar la variable, la probabilidad anterior es

$$1 - p\left[\frac{X-5}{0.5} \leq \frac{6-5}{0.5}\right] = 1 - \Phi\left(\frac{6-5}{0.5}\right) = 1 - \Phi(2) ,$$

donde $\Phi(2)$ es el valor de la función de distribución normal estándar en el punto 2. Buscando en las tablas de la distribución $N(0,1)$, obtenemos:

$$p[X > 6] = 1 - \Phi(2) = 1 - 0.97725 = 0.02275 .$$

Por tanto, hay un 2.27 % de tubos con diámetro superior a 6 centímetros. ◁

En muchas aplicaciones estadísticas interesa conocer el porcentaje de valores de la variable que difiere de la media en más de un número k de desviaciones ($k = 1, 2, 3$). Dada una variable X con distribución $N(\mu, \sigma)$, tenemos que hallar las probabilidades de los sucesos de la forma $[|X - \mu| \geq k\sigma]$. En concreto, si $k = 1$,

$$p[|X - \mu| > \sigma] = 1 - p[-\sigma \leq X - \mu \leq \sigma] = 1 - p\left[-1 \leq \frac{X-\mu}{\sigma} \leq 1\right] .$$

Ahora bien, $Z = \frac{X-\mu}{\sigma}$ es una variable normal tipificada y, consiguientemente,

$$p[|X - \mu| \geq \sigma] = 1 - [\Phi(1) - \Phi(-1)] = 1 - [\Phi(1) - (1 - \Phi(1))] = 2[1 - \Phi(1)] .$$

Consultando en las tablas el valor de $\Phi(1)$, 0.8413, obtenemos:

$$p[|X - \mu| > \sigma] = 0.3174 ,$$

o, lo que es lo mismo,

$$p[|X - \mu| \leq \sigma] = 0.6826 .$$

Es decir, el 31.74 % de valores de una normal $N(\mu, \sigma)$ difiere de la media en más de una desviación.

Haciendo el mismo razonamiento para $k = 2$,

$$p[|X - \mu| > 2\sigma] = 0.0455$$

o bien,

$$p[|X - \mu| \leq 2\sigma] = 0.9545 .$$

Esto es, el 4.55 % de valores de una normal difiere de la media en más de dos desviaciones.

Por último, para $k = 3$,

$$p[|X - \mu| > 3\sigma] = 0.00135$$

o

$$p[|X - \mu| \leq 3\sigma] = 0.99865 ,$$

o sea, únicamente el 0.135 % de valores de una distribución normal difiere de la media en más de tres desviaciones. Adviértase que, si bien la variable normal toma valores en toda la recta real, podemos decir que prácticamente todos, un 99.865 %, se encuentra en el intervalo $[\mu - 3\sigma, \mu + 3\sigma]$.

La figura 10.7 recoge todas estas probabilidades bajo la curva que representa la función de densidad normal.

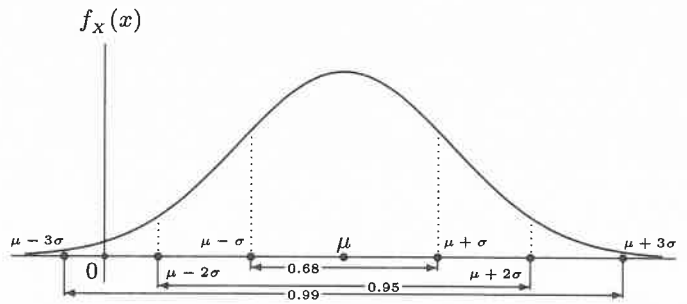


FIG. 10.7.

La siguiente proposición muestra que las combinaciones lineales de variables normales independientes son, asimismo, normales. En particular, la distribución normal es reproductiva, como indica el primer corolario que sigue a este resultado.

Proposición 10.5 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes, donde X_j tiene una distribución $N(\mu_j, \sigma_j)$, para $j = 1, \ldots, n$. Sean $a_1, \ldots, a_n$ y b números reales. Entonces, la variable $Y = a_1X_1 + \cdots + a_nX_n + b$ tiene una distribución $N\left(a_1\mu_1 + \cdots + a_n\mu_n + b,\ \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}\right)$.*

Demostración. Obtengamos la función característica de Y.

$$\begin{aligned}\varphi_Y(t) = E\left(e^{itY}\right) &= E[\exp\{it(a_1X_1 + \cdots + a_nX_n + b)\}] = \\ &= e^{itb} \cdot E(e^{ita_1X_1}) \cdots E(e^{ita_nX_n}) = e^{itb} \cdot \varphi_{X_1}(ta_1) \cdots \varphi_{X_n}(ta_n) = \\ &= e^{itb} \cdot \exp\left\{ita_1\mu - \frac{-t^2a_1^2\sigma_1^2}{2}\right\} \cdots \exp\left\{ita_n\mu - \frac{-t^2a_n^2\sigma_n^2}{2}\right\} = \\ &= \exp\left\{it\left(\sum_{j=1}^{n} a_j\mu_j + b\right) - \frac{t^2}{2}\sum_{j=1}^{n} a_j^2\sigma_j^2\right\}.\end{aligned}$$

Por tanto, la función característica de Y es la de una distribución normal con esperanza $a_1\mu_1 + \cdots + a_n\mu_n + b$ y varianza $a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$. En consecuencia, por el teorema de unicidad (corolario 8.6.1)

$$Y \hookrightarrow N\left(a_1\mu_1 + \cdots + a_n\mu_n + b, \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}\right). \clubsuit$$

Un importante resultado, que se deduce de la proposición anterior, es que la suma de variables aleatorias normales e independientes es una variable aleatoria normal.

Corolario 10.5.1 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes, donde X_j tiene una distribución $N(\mu_j, \sigma_j)$, para $j = 1, \ldots, n$. La variable $Y = X_1 + \cdots + X_n$ sigue una distribución $N\left(\mu_1 + \cdots + \mu_n, \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}\right)$.*

Demostración. Basta hacer $a_1 = \cdots = a_n = 1$ y $b = 0$, en la expresión obtenida en la proposición 10.5. $\clubsuit$

Corolario 10.5.2 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes e igualmente distribuidas $N(\mu, \sigma)$. La variable media muestral,*

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n},$$

sigue una distribución $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Demostración. Haciendo $a_1 = \cdots = a_n = \frac{1}{n}$ y $b = 0$ en la proposición anterior, resulta:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \hookrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right). \clubsuit$$

Corolario 10.5.3 *Sea X una variable aleatoria con distribución $N(\mu, \sigma)$. La variable $Y = a \cdot X + b$ tiene una distribución $N(a\mu + b, a\sigma)$.*

La demostración es inmediata a partir de la proposición, sin más que tomar una sola variable.

10.3 Distribución logarítmico-normal

Una variable aleatoria, X, sigue una **distribución logarítmico-normal** de parámetros μ y σ, si la variable $Y = \ln X$ sigue una distribución $N(\mu, \sigma)$. Se denota abreviadamente como

$$X \hookrightarrow \log N(\mu, \sigma) \, .$$

Calculemos su función de densidad a partir de su definición:

$$F_X(x) = p[X \leq x] = p[\ln X \leq \ln x] = p[Y \leq \ln x] = F_Y(\ln x) \, ,$$

con lo cual, derivando,

$$f_X(x) = f_Y(\ln x) \cdot \frac{1}{x} \, .$$

Sustituyendo en la expresión anterior la función de densidad de Y, obtenemos la **función de densidad** de la distribución logarítmico-normal, también llamada de COBB-DOUGLAS:

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}x\sigma} \cdot e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

En la figura 10.8 se representa esta función de densidad.

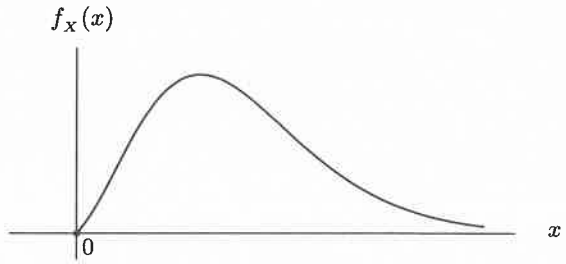


FIG. 10.8.

La variable logarítmico-normal o log-normal, que no es sino la exponencial de una variable aleatoria normal, fue propuesta por GALTON en 1879, estudiando el comportamiento de la media geométrica de n variables aleatorias independientes (como es sabido, el logaritmo neperiano de la media geométrica de n variables es la media aritmética de los neperianos de las variables).

A partir de 1930, una vez conocidas sus propiedades estadísticas, se multiplican sus aplicaciones: estudio de la magnitud de ciertas unidades económicas, modelización de las dosis críticas de ciertos medicamentos, tamaños de partículas, estudios literarios, agrícolas, zoológicos, etc.

La **función de distribución** puede expresarse en términos de la función de distribución de una normal tipificada:

$$F_X(x) = F_Y(\ln x) = p[Y \leq \ln x] = p\left(\frac{Y-\mu}{\sigma} \leq \frac{\ln x - \mu}{\sigma}\right) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) .$$

Por tanto, $F_X(x)$ es el valor de la función de distribución de la normal estándar en el punto $\frac{\ln x-\mu}{\sigma}$.

Ejemplo 10.5. Los ingresos mensuales, en miles de pesetas, de los individuos de un país siguen una distribución logarítmico-normal de parámetros $\mu = 3$ y $\sigma = 2$. ¿Cuál es el porcentaje de la población cuyos ingresos están comprendidos entre 70 y 150 mil pesetas?

Designando por X a la variable aleatoria *ingresos mensuales*, tendremos que hallar la probabilidad del suceso $[70 \leq X \leq 150]$ donde $X \hookrightarrow \log N(3,2)$. Entonces,

$$\begin{aligned} p[70 \leq X \leq 150] &= p[\ln(70) \leq \ln(X) \leq \ln(150)] = \\ &= p\left[\frac{\ln(70)-3}{2} \leq \frac{\ln(X)-3}{2} \leq \frac{\ln(150)-3}{2}\right] . \end{aligned}$$

Ahora bien, la variable aleatoria $\frac{\ln(X)-3}{2}$ sigue una distribución normal estándar, por lo cual,

$$p[70 \leq X \leq 150] = \Phi\left(\frac{\ln(150)-3}{2}\right) - \Phi\left(\frac{\ln(70)-3}{2}\right) = \Phi(1) - \Phi(0.62) = 0.1089 .$$

Es decir, el 10.89 % de los individuos tiene ingresos comprendidos entre 70 y 150 mil pesetas. ◁

Calculemos los principales momentos de la variable. Para obtener la expresión del momento de orden k, realizamos el cambio de variable $t = \frac{\ln x-\mu}{\sigma}$:

$$\begin{aligned} E(X^k) &= \int_{x=0}^{\infty} x^k \cdot \frac{1}{\sqrt{2\pi}x\sigma} \cdot e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}} dx = \\ &= \int_{t=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left(e^{t\sigma+\mu}\right)^k \cdot e^{-\frac{1}{2}t^2} dt = \\ &= \int_{t=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{k\mu} \cdot e^{\frac{-(t^2-2kt\sigma)}{2}} dt = \\ &= e^{(k\mu+\frac{k^2\sigma^2}{2})} \cdot \int_{t=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-(t-k\sigma)^2}{2}} dt = \\ &= e^{k\mu+\frac{k^2\sigma^2}{2}} , \end{aligned}$$

donde la última integral vale la unidad, pues se trata de la integral, en todo su campo de variación, de la función de densidad de una normal de media $k\sigma$ y desviación típica 1.

Concretando esta expresión para $k = 1$ y 2, se deduce la **esperanza**

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

y el momento central de orden 2,

$$E(X^2) = e^{2\mu + 2\sigma^2} .$$

Con lo cual, **la varianza** será

$$Var(X) = e^{2\mu} \cdot (e^{2\sigma^2} - e^{\sigma^2}) .$$

La siguiente proposición y sus corolarios muestran propiedades análogas a las obtenidas en la proposición 10.5 para la distribución normal, sustituyendo sumas por productos, y éstos por potencias. En concreto, el primero de los corolarios indica que la distribución logarítmico-normal es reproductiva para el producto, esto es, que el producto de variables logarítmico-normales e independientes sigue también una distribución logarítmico-normal.

Proposición 10.6 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes, donde X_i tiene una distribución $\log N(\mu_i, \sigma_i)$, para $i = 1, \ldots, n$. Sean $a_1, \ldots, a_n$ números reales. Entonces, $Y = X_1^{a_1} \cdots X_n^{a_n}$ sigue una distribución*

$$\log N\left(a_1\mu_1 + \cdots + a_n\mu_n, \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}\right) ,$$

Demostración. Tomando logaritmos neperianos,

$$\ln(Y) = \ln\left(\prod_{i=1}^{n} X_i^{a_i}\right) = \sum_{i=1}^{n} a_i \ln X_i .$$

Además, si $X_1, \ldots, X_n$ son variables aleatorias independientes con distribución logarítmico-normal $\log N(\mu_i, \sigma_i)$ $(i = 1, \ldots, n)$, entonces, $\ln X_1, \ldots, \ln X_n$ serán también independientes y con distribución $N(\mu_i, \sigma_i)$. Aplicando la proposición 10.5, la combinación lineal anterior será también normal; en concreto,

$$\ln(Y) = \ln\left(\prod_{i=1}^{n} X_i^{a_i}\right) = \sum_{i=1}^{n} a_i \ln X_i \hookrightarrow N\left(a_1\mu_1 + \cdots + a_n\mu_n, \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}\right) .$$

Sin más que considerar la definición de distribución logarítmico-normal,

$$Y = X_1^{a_1} \ldots X_n^{a_n} \hookrightarrow \log N\left(a_1\mu_1 + \cdots + a_n\mu_n, \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}\right) . \clubsuit$$

El siguiente corolario indica, como antes decíamos, que la variable log-normal es reproductiva para el producto.

Corolario 10.6.1 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes, donde X_i tiene una distribución $\log N(\mu_i, \sigma_i)$, para $i = 1, \ldots, n$. La variable $X_1 \cdots X_n$ sigue una distribución* $\log N\left(\mu_1 + \cdots + \mu_n, \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}\right)$.

Basta aplicar la proposición anterior para $a_1 = \cdots = a_n = 1$.

La variable log-normal tiene, para la media geométrica, un comportamiento similar al que demostrábamos para la media aritmética en el caso de distribuciones normales, como pone en evidencia el siguiente corolario.

Corolario 10.6.2 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes y tal que X_i tiene una distribución $\log N(\mu_i, \sigma_i)$, para $i = 1, \ldots, n$. La variable media geométrica,*

$$\overline{X}_G = \sqrt[1/n]{X_1 \cdots X_n}\,,$$

sigue una distribución $\log N\left(\frac{\mu_1 + \cdots + \mu_n}{n}, \frac{\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}}{n}\right)$.

Aplicando la proposición anterior para $a_1 = \cdots = a_n = \frac{1}{n}$, se deduce el resultado.

Obviamente, si $X_1, \ldots, X_n$ están igualmente distribuidas, $\log N(\mu, \sigma)$, su media geométrica, $\overline{X}_G$, tendrá una distribución $\log N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

10.4 Distribución gamma

Se dice que una variable aleatoria X sigue una **distribución gamma** de parámetros a y p (con $a, p \in \mathbb{R}$ y tales que $a > 0, p > 0$), si su **función de densidad** es

$$f_X(x) = \begin{cases} \frac{a^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

donde $\Gamma(p)$ es una de las funciones eulerianas gamma definida en la nota al pie número 2 de este mismo capítulo, $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$.

Cuando una variable tiene esta distribución, la denotaremos abreviadamente como

$$X \hookrightarrow \gamma(a, p)\,.$$

Comprobemos que la función de densidad está bien definida. Evidentemente, es positiva; comprobemos que la integral en todo su campo de variación es la unidad. Haciendo el cambio de variable $ax = t$, tendremos,

$$\begin{aligned} \int_{x=-\infty}^{\infty} f_X(x)dx &= \int_{x=0}^{\infty} \frac{a^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-ax} dx = \int_{t=0}^{\infty} \frac{a^{p-1}}{\Gamma(p)} \cdot \left(\frac{t}{a}\right)^{p-1} \cdot e^{-t} dt = \\ &= \frac{1}{\Gamma(p)} \cdot \int_{x=0}^{\infty} t^{p-1} \cdot e^{-t} dt = 1\,. \end{aligned}$$

En la figura 10.9 se representan distintas funciones de densidad de distribuciones gamma.

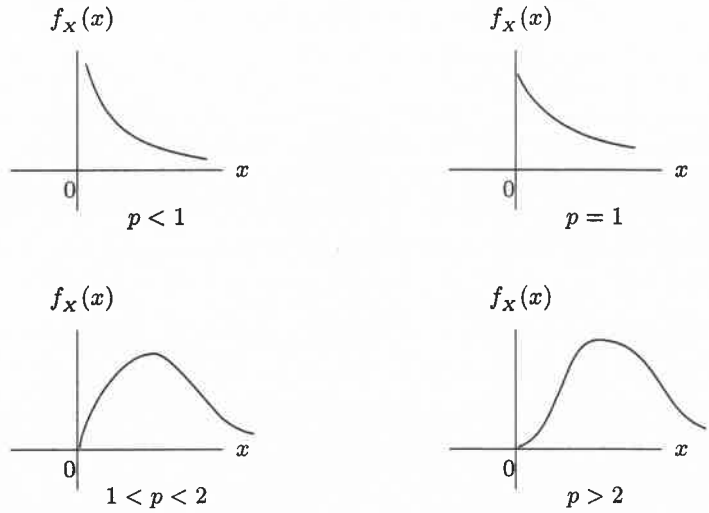


FIG. 10.9.

Las variables gamma aparecen en 1836 en los trabajos de P. S. DE LAPLACE, y a lo largo del pasado siglo proporcionan un gran número de aplicaciones, sobre todo las variables denominadas χ^2 (que no son sino variables gamma de un tipo particular), en el estudio de diversas medidas relacionadas con variables normales y multinomiales.

En nuestro siglo, su utilización se amplía a la modelización de variables aleatorias con valores no negativos, sustituyendo, a veces, a la distribución normal, y al estudio de los llamados tiempos de vida de componentes en el análisis de sistemas. En general, las variables del tipo *tiempo de espera* se ajustan bien a una distribución gamma, cuando el tiempo se mide de modo continuo.

Su **función de distribución** es

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \frac{a^p}{\Gamma(p)} \cdot t^{p-1} \cdot e^{-at} & x > 0 \end{cases}$$

Aunque se conoce la primitiva de esta integral cuando p es un número entero positivo (integrando por partes), en general sus probabilidades se calculan de forma aproximada. Existen tablas que recogen la función de distribución de distribuciones gamma para distintos valores de los parámetros[4].

4. Puede consultarse, por ejemplo, *Tables of The Incomplete Gamma Function*, editado por K. PEARSON, F.R.S., Cambridge, 1965.

Calculemos el momento de orden k de esta variable:

$$\begin{aligned} E(X^k) &= \int_0^\infty x^k \cdot \frac{a^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-ax} dx = \\ &= \frac{\Gamma(p+k)}{a^k \Gamma(p)} \cdot \int_0^\infty \frac{a^{p+k}}{\Gamma(p+k)} \cdot x^{p+k-1} \cdot e^{-ax} dx = \\ &= \frac{\Gamma(p+k)}{a^k \Gamma(p)} , \end{aligned}$$

donde la última integral vale la unidad, pues se trata de la integral, en todo el campo de variación, de la función de densidad de una distribución $\gamma(a, p+k)$.

Así, teniendo en cuenta algunas de las propiedades de la función Γ ya citadas, se deduce de modo inmediato que la **esperanza** vale

$$E(X) = \frac{\Gamma(p+1)}{a\Gamma(p)} = \frac{p}{a} .$$

Además,

$$E(X^2) = \frac{p^2+p}{a^2}$$

y, por tanto, la **varianza** vale

$$Var(X) = \frac{p}{a^2} .$$

El cálculo de la **función característica** conlleva la resolución de una integral compleja, por lo que desborda el nivel de este libro. Con todo, ilustremos el resultado con una demostración no formal.

$$\begin{aligned} \varphi_X(t) &= \int_0^\infty e^{itx} \cdot \frac{a^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-ax} dx = \\ &= \frac{a^p}{(a-it)^p} \cdot \int_0^\infty \frac{(a-it)^p}{\Gamma(p)} \cdot x^{p-1} \cdot e^{-(a-it)x} dx = \\ &= \frac{a^p}{(a-it)^p} = \left(1 - \frac{it}{a}\right)^{-p} . \end{aligned}$$

donde, de nuevo, el lector debe aceptar que la última integral vale uno, por analogía con las integrales entre 0 e ∞ de las funciones de densidad de las variables gamma.

Ejemplo 10.6. La empresa de Juan produce un cierto tipo de piezas y, una vez vendidas, se hace cargo del coste de la primera reparación. El director del departamento de ventas de la empresa ha ordenado a Juan que determine la cantidad que ha de costar la hora de reparación de dichas piezas para que, siendo el coste de fabricación mil pesetas, el precio de venta 5 mil y la distribución del tiempo de reparación una gamma de parámetros $a = 0.5$ y $p = 2$, el beneficio medio sea de 3 mil pesetas.

Sabemos que el beneficio es la diferencia entre ingresos y gastos. En este caso, el ingreso es fijo (5 mil pesetas) y los gastos mil pesetas fijas de fabricación y $C \cdot X$ de reparación, donde X es la variable que designa el tiempo, en horas, empleado en la

reparación, y C el coste por hora de la misma. Por tanto, el beneficio en miles de pesetas, B, es

$$B = 5 - (CX + 1)$$

y, consecuentemente, el beneficio medio es

$$E(B) = 5 - [C \cdot E(X) + 1] .$$

Como el tiempo medio que se tarda en reparar una pieza es

$$E(X) = \frac{2}{0.5} = 4 \text{ horas} ,$$

y se desea que el beneficio medio sea de 3 mil pesetas, bastará despejar para obtener $C = 0.25$. Es decir, hay que cobrar 250 pesetas por hora de reparación, si se espera obtener un beneficio de 3 mil pesetas por pieza. ◁

En la siguiente proposición se demuestra que la distribución gamma es reproductiva, siempre que el parámetro a sea el mismo para todos los sumandos, esto es, la suma de variables gamma independientes con el mismo parámetro a es una variable gamma.

Proposición 10.7 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes y tales que X_j tiene una distribución $\gamma(a, p_j)$, para $j = 1, \ldots, n$. Entonces, la variable $Y = X_1 + \cdots + X_n$ sigue una distribución $\gamma(a, p_1 + \cdots + p_n)$.*

Demostración. Calculando la función característica de Y:

$$\varphi_Y(t) = \varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) = \prod_{j=1}^n \left(1 - \frac{it}{a}\right)^{-p_j} = \left(1 - \frac{it}{a}\right)^{-\sum_{j=1}^n p_j}$$

obtenemos la de una distribución $\gamma(a, p_1 + \cdots + p_n)$. El teorema de unicidad nos dice que, entonces, $Y = \sum_{j=1}^n X_j$ sigue esta distribución. ♣

Una última proposición nos informa sobre el efecto que produce sobre una variable gamma el multiplicarla por una constante positiva.

Proposición 10.8 *Sea X una variable aleatoria con distribución $\gamma(a, p)$. Entonces, la variable $k \cdot X$, donde $k > 0$, sigue una distribución $\gamma\left(\frac{a}{k}, p\right)$.*

Demostración.

$$\varphi_{kX}(t) = \varphi_X(kt) = \left(1 - \frac{ikt}{a}\right)^{-p} = \left(1 - \frac{it}{a/k}\right)^{-p} ,$$

de donde se obtiene el resultado. ♣

Corolario 10.8.1 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes e igualmente distribuidas con distribución $\gamma(a,p)$. La variable media muestral*

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} ,$$

sigue una distribución $\gamma(an, pn)$.

Demostración. De la proposición 10.7 se deduce que la suma, $\sum_{i=1}^{n} X_i$ sigue una distribución $\gamma(a, pn)$. Ahora bien, como $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, se demuestra el resultado, sin más que aplicar la proposición anterior para $k = \frac{1}{n}$. ♣

10.5 Distribución exponencial negativa

Se dice que una variable aleatoria X sigue una **distribución exponencial negativa** de parámetro a ($a \in \mathbb{R}$, $a > 0$), si su **función de densidad** es

$$f_X(x) = \begin{cases} a \cdot e^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

En la figura 10.10 se representa esta función de densidad. Denotaremos abreviadamente esta variable como

$$X \hookrightarrow \mathcal{E}(a) .$$

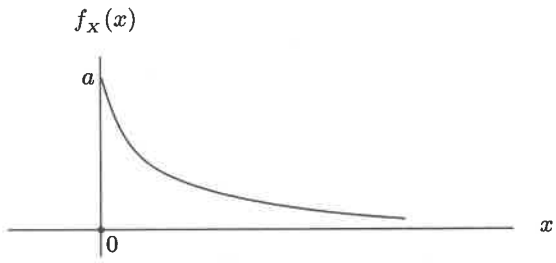


FIG. 10.10.

Puede comprobar el lector, comparando las funciones de densidad, que la distribución exponencial negativa, también denominada simplemente distribución exponencial, es un caso particular de la distribución gamma cuando el parámetro p vale uno, es decir, $\mathcal{E}(a) \equiv \gamma(a, 1)$. Nos basaremos en este hecho para hallar sus momentos y su función característica.

La variable exponencial negativa aparece de forma natural en el estudio de los tiempos de vida de ciertos procesos (es decir, los denominados tiempos de espera). En este sentido, juega el mismo papel que la distribución geométrica (discreta), pero cuando el tiempo se mide de manera continua.

Integrando la función de densidad, deducimos la **función de distribución**:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-ax} & x > 0 \end{cases}$$

que se representa en la figura 10.11.

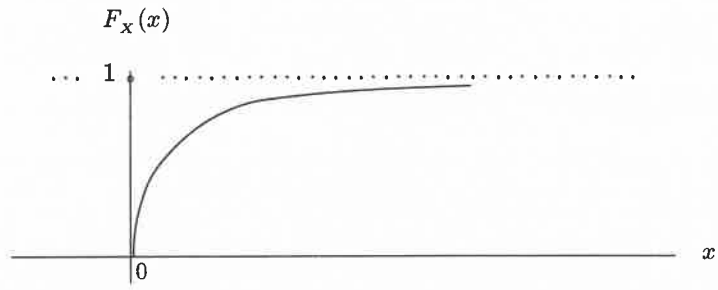


FIG. 10.11.

Obsérvese que se obtiene una expresión explícita de la misma, expresión que, además, es sencilla de hallar para los distintos valores x. El cálculo de probabilidades para una distribución exponencial no presenta, por tanto, problemas.

Los resultados obtenidos para la distribución gamma permiten calcular los momentos de la variable exponencial, sin más que concretar para $p = 1$. Así,

$$E(X^k) = \frac{\Gamma(1+k)}{a^k \Gamma(1)} = \frac{\Gamma(1+k)}{a^k} ;$$

en consecuencia, la **esperanza** será

$$E(X) = \frac{1}{a}$$

y la **varianza**,

$$Var(X) = \frac{1}{a^2} .$$

De igual forma, su **función característica** es

$$\varphi_X(t) = \left(1 - \frac{it}{a}\right)^{-1} .$$

Proposición 10.9 *Sean $X_1, \ldots, X_n$ variables aleatorias independientes e igualmente distribuidas, donde X_i tiene distribución $\mathcal{E}(a)$. Entonces, $Y = X_1 + \cdots + X_n$ sigue una distribución $\gamma(a, n)$.*

Se deduce de inmediato de la proposición 10.7, considerando la exponencial como una gamma.

En consecuencia, la distribución exponencial no es reproductiva, aunque la distribución de la suma de exponenciales independientes es conocida, lo que facilita su uso en las aplicaciones de la Inferencia Estadística.

La distribución exponencial cumple una propiedad que ya vimos para la distribución geométrica: la propiedad del *olvido* o de la *falta de memoria.*

Proposición 10.10 *Sea X una variable aleatoria con distribución $\mathcal{E}(a)$. Entonces,*

$$p[X > s+t \mid X > s] = p[X > t] \qquad \text{con } s,t \in \mathbb{R}\,,\ s,t > 0\,.$$

Demostración. Como

$$p[X > x] = 1 - p[X \leq x] = 1 - F_X(x) = e^{-ax}\,,$$

tendremos, tras aplicar la definición de probabilidad condicionada, que

$$\begin{aligned} p[X > s+t \mid X > s] &= \frac{p[X > s+t, X > s]}{p[X > s]} = \frac{p[X > s+t]}{p[X > s]} = \\ &= \frac{e^{-a(s+t)}}{e^{-as}} = e^{-at} = p[X > t]\,. \clubsuit \end{aligned}$$

Proposición 10.11 *Una variable aleatoria continua que tome valores positivos y verifique la propiedad del olvido o falta de memoria, es necesariamente una variable aleatoria exponencial negativa.*

Demostración. De la propiedad del olvido se deduce que

$$p[X > s+t] = p[X > s] \cdot p[X > t]\,,$$

de donde, sin más que tener en cuenta la expresión de la probabilidad de un suceso complementario, obtenemos

$$1 - F_X(s+t) = [1 - F_X(s)] \cdot [1 - F_X(t)]$$

o, equivalentemente,

$$F_X(s+t) = F_X(s) + F_X(t) - F_X(s) \cdot F_X(t)\,.$$

Derivando respecto de s,

$$f_X(s+t) = f_X(s) \cdot [1 - F_X(t)]\,,$$

y haciendo $s = 0$, resulta:

$$\frac{\partial F_X(t)}{\partial t} = f_X(t) = a \cdot [1 - F_X(t)]\,,$$

con $a = f_X(0)$. La resolución de esta ecuación diferencial nos conduce a $1 - F_X(t)$ proporcional a e^{-at}. Por tanto,

$$F_X(t) = 1 - e^{-at}\,,$$

expresión de la función de distribución de una variable aleatoria exponencial negativa de parámetro a. ♣

Las dos proposiciones anteriores son análogas, para variables continuas, al resultado que probamos en el capítulo anterior (proposición 9.8) para la distribución geométrica. De ellas deducimos que la propiedad del olvido o de la falta de memoria caracteriza a la exponencial entre las distribuciones continuas y no negativas.

Si la variable X representa el tiempo de vida de un equipo, esto es, el tiempo que tarda en sufrir una avería, la propiedad de la falta de memoria dice que, si lleva funcionando un tiempo s, la probabilidad de que continúe activo un tiempo adicional t sólo depende de t, por lo que coincide con la probabilidad de que una pieza nueva no falle durante un tiempo t. Esto significa que los modelos exponenciales deben aplicarse al estudio del *tiempo de vida* de piezas que no sufran desgaste o deterioro en su funcionamiento.

Ejemplo 10.7. El tiempo, en días, que tarda la empresa de Juan en servir los pedidos a sus clientes es una variable aleatoria con distribución exponencial negativa de media 5 días. Han pasado 2 días y un cliente aún no ha recibido su pedido. ¿Cuál es la probabilidad de que tenga que esperar al menos 2 días más?

Hemos de calcular

$$p[X > 4 \mid X > 2] ,$$

donde X es la variable con distribución $\mathcal{E}\left(\frac{1}{5}\right)$.

La propiedad del olvido nos dice que dicha probabilidad es

$$p[X > 4 \mid X > 2] = p[X > 2] = e^{-2/5} = 0.670 . \triangleleft$$

10.6 Distribución beta

Una variable aleatoria X sigue una **distribución beta** de parámetros p y q ($p, q \in \mathbb{R}$, $p > 0, q > 0$), si su **función de densidad** es

$$f_X(x) = \begin{cases} \frac{1}{\beta(p,q)} \cdot x^{p-1} \cdot (1-x)^{q-1} & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

donde $\beta(p, q)$ es la función euleriana beta:

$$\beta(p, q) = \int_0^1 x^{p-1} \cdot (1-x)^{q-1} dx .$$

El cálculo de la función beta para distintos valores de p y q se realiza sin dificultad, si recordamos que

$$\beta(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)} .$$

Esta expresión permite escribir la función de densidad en la forma

$$f_X(x) = \begin{cases} \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} \cdot x^{p-1} \cdot (1-x)^{q-1} & 0 < x < 1 \\ 0 & \text{en el resto} \end{cases}$$

Si X es una variable beta de parámetros p y q, lo denotamos abreviadamente como

$$X \hookrightarrow \beta(p,q) .$$

La función de densidad de una beta está bien definida. En efecto, es positiva y la integral en todo el campo es

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)dx &= \int_0^1 \frac{1}{\beta(p,q)} \cdot x^{p-1} \cdot (1-x)^{q-1} dx = \\ &= \frac{1}{\beta(p,q)} \cdot \int_0^1 x^{p-1} \cdot (1-x)^{q-1} dx = \\ &= \frac{1}{\beta(p,q)} \cdot \beta(p,q) = 1 .\end{aligned}$$

El aspecto de la función de densidad es muy diferente según que los valores de p y q sean unos u otros. Podemos conseguir, eligiéndolos adecuadamente, que la densidad sea creciente, decreciente, con aspecto acampanado, en U, etc. Como muestra, la figura 10.12 recoge cuatro tipos de funciones de densidad de una distribución beta.

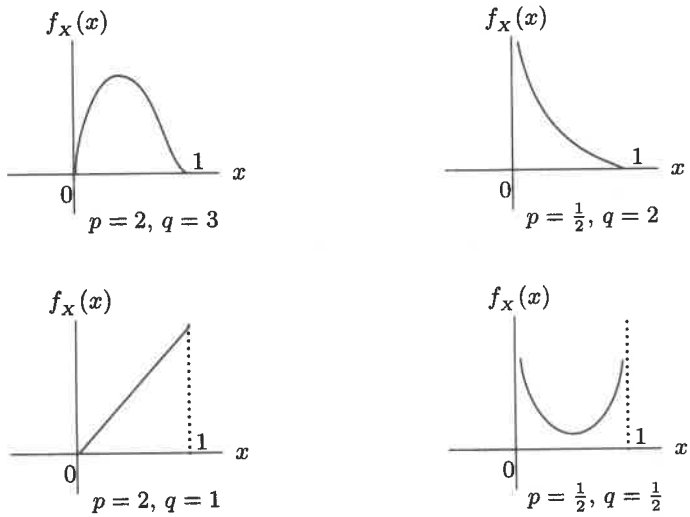


FIG. 10.12.

Dada la gran flexibilidad de la distribución beta y el hecho de que tome valores entre cero y uno, las variables que describen proporciones (proporción de riqueza en un mineral, incidencia del presupuesto para alimentación en el presupuesto familiar, proporción de manufacturas defectuosas, etc.), se ajustan bien a dicha distribución, si sabemos encontrar los valores más adecuados de los parámetros.

Su **función de distribución** vale

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \frac{\Gamma(p+q)}{\Gamma(p)\cdot\Gamma(q)} t^{p-1}(1-t)^{q-1}dt & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Para el cálculo de valores de la función de distribución (y, por tanto, para el de las probabilidades), se utilizan tablas que los proporcionan con una cierta aproximación[5].

Hallemos los momentos. El momento de orden k, $E(X^k)$, de esta variable es

$$E(X^k) = \int_0^1 x^k \cdot \frac{\Gamma(p+q)}{\Gamma(p)\cdot\Gamma(q)} x^{p-1}(1-x)^{q-1}dx \ ,$$

o, lo que es lo mismo,

$$\frac{\Gamma(p+q)\cdot\Gamma(p+k)}{\Gamma(p)\cdot\Gamma(p+q+k)} \int_0^1 \frac{\Gamma(p+k+q)}{\Gamma(p+k)\cdot\Gamma(q)} \cdot x^{p+k-1} \cdot (1-x)^{q-1}dx \ ;$$

puesto que la integral anterior vale uno por serlo de la función de densidad de una distribución $\beta(p+k, q)$, en todo el campo de definición, obtenemos que

$$E(X^k) = \frac{\Gamma(p+q)\cdot\Gamma(p+k)}{\Gamma(p)\cdot\Gamma(p+q+k)} \ .$$

Haciendo $k = 1$ y $k = 2$ en el resultado anterior, tendremos los momentos de primer y segundo orden; así, la **esperanza**

$$E(X) = \frac{p}{p+q}$$

y

$$E(X^2) = \frac{p(p+1)}{(p+q+1)(p+q)} \ .$$

Deducimos, entonces, que la **varianza** vale

$$Var(X) = \frac{pq}{(p+q+1)(p+q)^2} \ .$$

Ejemplo 10.8. La proporción media anual de tubos de acero defectuosos en el proceso de fabricación es 0.2. Sabiendo que dicha proporción se modeliza según una distribución beta cuyo parámetro q es igual a 8, determinemos el valor del parámetro p.

Aplicando el valor de la esperanza, tendremos:

$$E(X) = 0.2 = \frac{p}{p+8} \ .$$

Con lo cual, sin más que despejar, deducimos que $p = 2$. ◁

La siguiente proposición permite calcular probabilidades para una distribución $\beta(p, q)$, cuando p y q son enteros.

5. El lector interesado en ello puede consultar *Tables of The Incomplete Beta Function*, editado por K. PEARSON, F.R.S., Cambridge, 1968.

Proposición 10.12 *Sea X una variable aleatoria con distribución $\beta(p,q)$ y sea a un número entre cero y uno. Entonces,*

$$p[X \leq a] = p[Y \geq p] \ ,$$

siendo Y una variable binomial de parámetros $p+q-1$ y a.

Demostración.

$$\begin{aligned} p[X \leq a] &= \int_0^a \frac{\Gamma(p+q)}{\Gamma(p)\cdot\Gamma(q)} x^{p-1}(1-x)^{q-1}dx = \\ &= \frac{(p+q-1)!}{(p-1)!\cdot(q-1)!}\int_0^a x^{p-1}(1-x)^{q-1}dx \ . \end{aligned}$$

Integrando por partes, obtenemos que la integral anterior vale

$$\frac{1}{p}a^p(1-a)^{q-1} + \frac{q-1}{p}\int_0^a x^p(1-x)^{q-2}dx \ .$$

Si reiteramos el procedimiento de integración $q-2$ veces, resulta:

$$\begin{aligned} \int_0^a x^{p-1}(1-x)^{q-1}dx &= \frac{1}{p}a^p(1-a)^{q-1} + \frac{q-1}{p(p+1)}a^{p+1}(1-a)^{q-2} + \\ &+ \frac{(q-1)(q-2)}{p(p+1)(p+2)}a^{p+2}(1-a)^{q-3} + \cdots + \\ &+ \frac{(q-1)(q-2)\cdots 1}{p(p+1)\cdots(p+q-1)}a^{p+q-1}(1-a)^0 \ . \end{aligned}$$

Por tanto, $p[X \leq a]$ se expresa como

$$\begin{aligned} p[X \leq a] &= \frac{(p+q-1)!}{(p-1)!\cdot(q-1)!}\left[\frac{1}{p}a^p(1-a)^{q-1} + \frac{q-1}{p(p+1)}a^{p+1}(1-a)^{q-2} + \cdots + \right. \\ &\left. + \frac{(q-1)(q-2)\cdots 1}{p(p+1)\cdots(p+q-1)}a^{p+q-1}(1-a)^0\right] = \\ &= \frac{(p+q-1)!}{p!\cdot(q-1)!}a^p(1-a)^{q-1} + \frac{(p+q-1)!}{(p+1)!\cdot(q-2)!}a^{p+1}(1-a)^{q-2} + \cdots + \\ &+ \frac{(p+q-1)!}{(p+q-1)!\cdot 0!}a^{p+q-1}(1-a)^0 = \\ &= \sum_{j=p}^{p+q-1}\binom{p+q-1}{j}\cdot a^j\cdot(1-a)^{p+q-1-j} = \\ &= \sum_{j=p}^{p+q-1} p[Y=j] = p[Y \geq p] \ , \end{aligned}$$

siendo Y una variable aleatoria con distribución $\mathcal{B}(p+q-1,a)$. ♣

10.7 Distribución de Pareto

Ante la necesidad de dar una explicación a la distribución de las rentas de los individuos de una población, el profesor suizo de economía V. PARETO planteó a finales del siglo pasado una distribución continua que cumpliera las dos condiciones siguientes:

1) No hay rentas inferiores a un cierto valor x_0 (renta mínima).

2) El porcentaje de individuos con una renta mayor o igual que un cierto valor decrece potencialmente al crecer dicho valor.

Es decir,

$$p[X \geq x] = A \cdot x^{-\alpha} , \qquad \text{para } x \geq x_0 ,$$

donde A es una constante que hay que determinar y α un parámetro positivo. Estas condiciones, como ahora veremos, conducen a una única distribución continua: la llamada **distribución de Pareto.**

Para calcular A basta tener en cuenta que la probabilidad del suceso $[X \geq x_0]$ ha de ser la unidad, es decir, $A \cdot x_0^{-\alpha} = 1$, con lo cual

$$A = x_0^{\alpha} .$$

Por lo que resulta:

$$p[X \geq x] = \left(\frac{x_0}{x}\right)^{\alpha} , \qquad \text{para } x \geq x_0 .$$

Durante mucho tiempo, esta distribución fue considerada la ley universal que explicaba la distribución de las rentas, válida para todas las condiciones políticas y sociales, como PARETO había sugerido. Aunque, como señala A. C. PIGOU, tal universalidad sea discutible, su uso se ha extendido en el estudio de muchos otros fenómenos empíricos, como el tamaño de las ciudades, las fluctuaciones en el precio de los stocks, etc.

La **función de distribución** de una variable X con distribución de Pareto de parámetros α y x_0 $(\alpha, x_0 \in \mathbb{R}, \alpha > 0, x_0 > 0)$ es

$$F(x) = p[X \leq x] = 1 - p[X \geq x] = \begin{cases} 0 & x < x_0 \\ 1 - (\frac{x_0}{x})^{\alpha} & x \geq x_0 \end{cases}$$

cuya representación gráfica aparece en la figura 10.13.

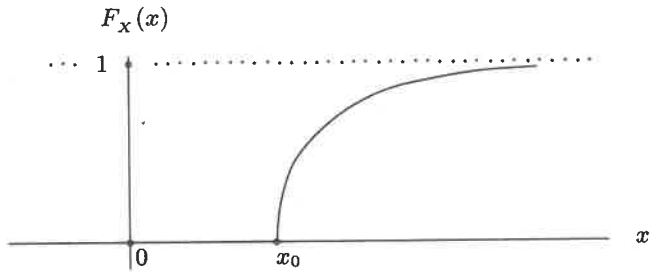


FIG. 10.13.

Derivando esta función, obtendremos la **función de densidad**:

$$f_X(x) = \begin{cases} 0 & x < x_0 \\ \frac{\alpha}{x} \cdot (\frac{x_0}{x})^\alpha & x \geq x_0 \end{cases}$$

cuya gráfica es la representada en la figura 10.14.

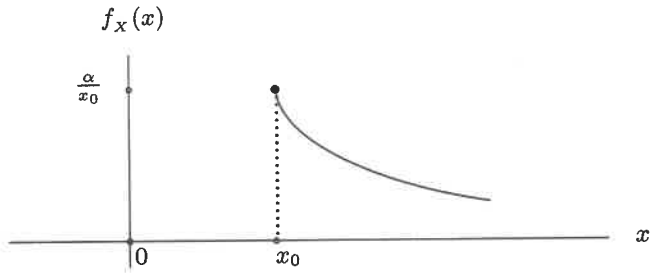


FIG. 10.14.

Esta función de densidad está bien definida. Por un lado, $f_X(x) \geq 0$ para todo x. Por otro,

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{x_0}^{\infty} \frac{\alpha}{x} \left(\frac{x_0}{x}\right)^\alpha dx = \left(-\frac{x_0}{x}\right)^\alpha \Big|_{x_0}^{\infty} = 1 .$$

Ejemplo 10.9. Los ingresos mensuales, en miles de pesetas, de los trabajadores en la empresa de Juan, siguen una distribución de Pareto con $x_0 = 50$ y $\alpha = 1.75$. ¿Qué porcentaje de individuos tiene un salario superior a 200 mil pesetas?

Basta calcular $p[X \geq 200]$ donde X, *salario mensual de los trabajadores*, es una variable con distribución de Pareto.

$$p[X \geq 200] = \left(\frac{50}{200}\right)^{1.75} = 0.088 .$$

Así, podemos decir que únicamente el 8.8 % de los trabajadores de la empresa percibe un salario mensual superior a 200 mil pesetas. ◁

Hallemos la expresión general de los momentos no centrales:

$$E(X^k) = \int_{x_0}^{\infty} x^k \cdot \frac{\alpha}{x} \cdot \left(\frac{x_0}{x}\right)^{\alpha} dx =$$
$$= \frac{\alpha x_0^k}{\alpha - k} \int_{x_0}^{\infty} \frac{\alpha - k}{x} \cdot \left(\frac{x_0}{x}\right)^{\alpha - k} dx = \frac{\alpha x_0^k}{\alpha - k}, \qquad \text{para } \alpha > k .$$

De nuevo, la última integral vale uno por ser la integral de la función de densidad de una variable de Pareto con parámetros x_0 y $\alpha - k$.

Para $k = 1$, obtenemos la **esperanza**:

$$E(X) = \frac{\alpha x_0}{\alpha - 1}, \quad \text{si } \alpha > 1 .$$

Cuando $\alpha \leq 1$, dicha esperanza no existe.

El momento de orden dos es

$$E(X^2) = \frac{\alpha x_0^2}{\alpha - 2},$$

que sólo existe para valores de α mayores que 2. De los dos momentos anteriores, se deduce la expresión de la **varianza**:

$$Var(X) = \frac{\alpha x_0^2}{(\alpha - 2)(\alpha - 1)^2}, \quad \text{para } \alpha > 2 .$$

10.8 Distribución de Cauchy

Una variable aleatoria X sigue una **distribución de Cauchy** de parámetros α y β ($\alpha, \beta \in \mathbb{R}$, $-\infty < \alpha < \infty$ y $\beta > 0$), si su **función de densidad** es

$$f_X(x) = \frac{1}{\beta \cdot \pi[1 + (\frac{x-\alpha}{\beta})^2]}, \quad \text{para } -\infty < x < \infty .$$

La función de densidad está bien definida. Evidentemente, es positiva y su integral en todo el campo de variación vale

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\beta \cdot \pi[1 + (\frac{x-\alpha}{\beta})^2]} dx =$$
$$= \frac{1}{\pi} \operatorname{arctg}\left(\frac{x-\alpha}{\beta}\right)\Big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} + \frac{\pi}{2}\right) = 1 .$$

Centraremos nuestra atención en la distribución de Cauchy de parámetros $\alpha = 0$ y $\beta = 1$, a la que llamaremos distribución de Cauchy estándar, cuya función de densidad es

$$f_X(x) = \frac{1}{\pi(1 + x^2)}, \qquad \text{para } -\infty < x < \infty .$$

Su aspecto gráfico recuerda al de la función de densidad $N(0,1)$, aunque decrece más lentamente cuando nos separamos del origen. Veamos sus propiedades, que el lector puede tratar de comprobar.

1) $f_X(x)$ es continua en todo $\mathbb{R}$.

2) $f_X(x)$ es simétrica respecto del origen, esto es, $f_X(x) = f_X(-x)$.

3) $f_X(x)$ nunca vale cero, pero tiene una asíntota horizontal en el eje de abscisas:

$$\begin{aligned} \lim_{x\to\infty} f_X(x) &= 0\,, \\ \lim_{x\to-\infty} f_X(x) &= 0\,. \end{aligned}$$

4) $f_X(x)$ es estrictamente creciente para $x < 0$ y estrictamente decreciente para $x > 0$. Para comprobarlo, basta observar la expresión de su derivada,

$$f'_X(x) = \frac{-2x}{\pi(1+x^2)^2}\,.$$

5) $f_X(x)$ presenta un máximo absoluto en el punto $\left(0, \frac{1}{\pi}\right)$.

6) $f_X(x)$ tiene dos puntos de inflexión, en $x = \frac{1}{\sqrt{3}}$ y en $x = \frac{-1}{\sqrt{3}}$.

A la vista de las anteriores propiedades, podemos dibujar dicha función de densidad como se muestra en la figura 10.15.

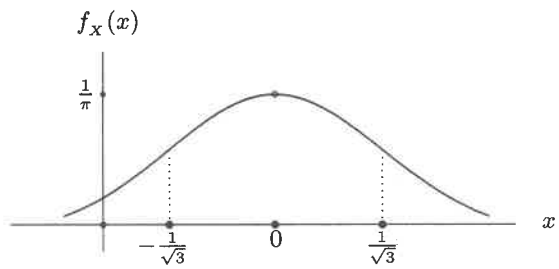


FIG. 10.15.

A diferencia de lo que ocurre con la distribución normal, su función de distribución admite una expresión explícita. En efecto,

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)}dt = \frac{\text{arctg}\,(x)}{\pi} + \frac{1}{2}\,, \qquad \text{cuando } -\infty < x < \infty\,.$$

Como se probó en el ejemplo 6.8, para una distribución de Cauchy no existe la esperanza y, consecuentemente, tampoco existen los momentos de otro tipo y orden.

Se comprueba (no lo demostraremos en este libro) que la **función característica** de una distribución de Cauchy vale:

$$\varphi_X(t) = e^{i\alpha t - \beta|t|} ,$$

concretándose para la distribución estándar en

$$\varphi_X(t) = e^{-|t|} .$$

Proposición 10.13 *Sean $X_1, \ldots, X_n$, n variables independientes e igualmente distribuidas con distribución de Cauchy estándar. La variable media muestral, $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$, también sigue una distribución de Cauchy estándar.*

Demostración. La función característica de la media muestral es:

$$\varphi_{\overline{X}}(t) = \varphi_{\sum_{j=1}^n X_j}\left(\frac{t}{n}\right) = \left[\varphi_{X_j}\left(\frac{t}{n}\right)\right]^n = \left(e^{-|t/n|}\right)^n = e^{-|t|} ,$$

luego $\overline{X}$ sigue una distribución de Cauchy. ♣

Ejemplo 10.10. Hallemos el intervalo intercuartílico de una distribución de Cauchy estándar, es decir,

$$R_c = C_3 - C_1 ,$$

donde C_1 y C_3 son los cuartiles de la distribución. Calculemos el tercer cuartil, teniendo en cuenta que

$$F_X(C_3) = 0.75 ,$$

siendo F_X la función de distribución de Cauchy estándar. Entonces,

$$\frac{\text{arctg}\,(C_3)}{\pi} + \frac{1}{2} = 0.75 ,$$

o bien,

$$\text{arctg}\,(C_3) = \frac{\pi}{4} .$$

Como la tangente de $\frac{\pi}{4}$ es la unidad,

$$C_3 = 1 .$$

Obviamente, $C_1 = -1$ sin más que considerar la simetría de la distribución de Cauchy. Por tanto, el intervalo intercuartílico será $[-1, 1]$, en el cual se encuentra el 50 % central de los valores de la variable.

En la distribución normal estándar el intervalo intercuartílico es $[-0.67, 0.67]$, que se obtiene buscando en las tablas los valores de C_1 y C_3 tales que

$$\Phi(C_1) = 0.25 \qquad \text{y} \qquad \Phi(C_3) = 0.75 .$$

En consecuencia, el 50 % central de los valores está más concentrado en la distribución $N(0,1)$ que en la distribución de Cauchy estándar. ◁

10.9 Distribución logística

La curva logística es una función real de variable real. En las aplicaciones conocidas, la variable que ocupa el eje de abscisas es el tiempo. La expresión analítica de la curva logística es

$$S(t) = \frac{a + bde^{\frac{t}{c}}}{1 + de^{\frac{t}{c}}} ,$$

donde a, b, c y d son constantes positivas, con $a < b$. Es una curva creciente, que tiende a su valor *final*, b, cuando t tiende a infinito, y a su valor *inicial*, a, cuando t tiende a menos infinito. Su derivada,

$$\frac{dS(t)}{dt} = \frac{c}{b-a}[S(t) - a] \cdot [b - S(t)] ,$$

indica que puede utilizarse para estudiar la variación en el tiempo de magnitudes cuya tasa de crecimiento en un instante, t, dependa multiplicativamente del exceso sobre el valor inicial, $[S(t) - a]$, y del defecto bajo el valor final, $[b - S(t)]$.

La curva logística aparece descrita por primera vez en 1845, dentro de un trabajo de P. F. VERHULST referente a las leyes que gobiernan el crecimiento de la población. Su uso se popularizó en la segunda mitad del siglo XIX. Se conocen numerosas aplicaciones dentro de nuestro siglo al estudio de la distribución de la renta, de la mortalidad, del crecimiento de la población, del tamaño del parque automovilístico, entre otras. En química, se ha utilizado para modelizar reacciones en las que uno de los productos que se combina sirve, a la vez, como catalizador (autocatálisis).

Cuando se estudian fenómenos aleatorios, la curva logística se emplea para representar el crecimiento de las probabilidades, esto es, para representar la función de distribución. Como las asíntotas de la curva logística anterior son a y b, se toma $a = 0$ y $b = 1$; asimismo, se fija $d = 1$.

Diremos que una variable aleatoria X sigue una **distribución logística** de parámetros α y β ($\alpha, \beta \in \mathbb{R}, -\infty < \alpha < \infty, \beta > 0$), si su **función de distribución** es

$$F_X(x) = \frac{e^{\frac{x-\alpha}{\beta}}}{1 + e^{\frac{x-\alpha}{\beta}}} = \frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}} , \quad \text{para } -\infty < x < \infty .$$

Obsérvese que, como antes indicamos, la función de distribución definida es una curva logística, con $a = 0$ y $b = d = 1$. Hemos escrito β en lugar de c y restado a la variable una cantidad α (un cambio de origen) para permitir que la distribución de probabilidades no esté centrada en el origen. Su representación gráfica aparece en la figura 10.16.

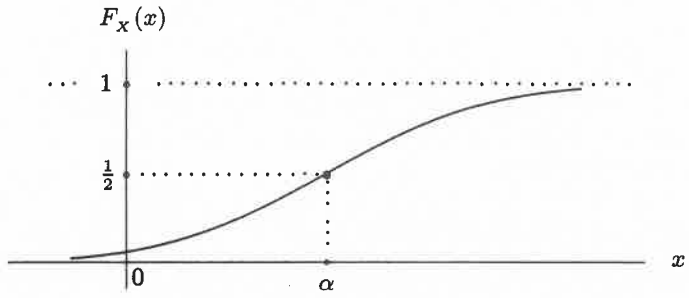


FIG. 10.16.

Derivando la función de distribución obtenemos su **función de densidad**,

$$f_X(x) = \frac{e^{-\frac{x-\alpha}{\beta}}}{\beta \cdot \left(1 + e^{-\frac{x-\alpha}{\beta}}\right)^2} \qquad -\infty < x < \infty ,$$

pudiéndose expresar, también, en términos de la secante hiperbólica (sech),

$$f_X(x) = \frac{1}{4\beta}\left[\operatorname{sech}\left(\frac{x-\alpha}{\beta}\right)\right]^2 .$$

La figura 10.17 recoge la representación gráfica de esta función que, como puede comprobar el lector, es simétrica respecto de α.

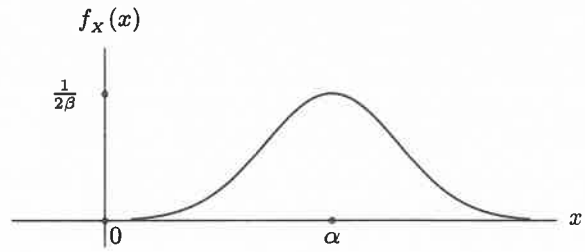


FIG. 10.17.

La función de densidad anterior está bien definida, pues toma valores positivos y

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{e^{-\frac{x-\alpha}{\beta}}}{\beta \cdot \left(1 + e^{-\frac{x-\alpha}{\beta}}\right)^2} dx = \left.\frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}}\right|_{x=-\infty}^{\infty} = 1 .$$

La versión estándar de la distribución logística resulta del cambio de variable $Y = \frac{X-\alpha}{\beta}$ (o lo que es lo mismo, tomando $\alpha = 0$ y $\beta = 1$), siendo su función de densidad

$$f_Y(y) = \frac{e^{-y}}{(1+e^{-y})^2}\,, \qquad -\infty < y < \infty$$

y su función de distribución

$$F_Y(y) = \frac{1}{1+e^{-y}}\,, \qquad -\infty < y < \infty\,.$$

Proporcionamos, a continuación, los principales momentos de una variable aleatoria logística, cuyos cálculos omitiremos. La **esperanza** vale

$$E(X) = \alpha\,,$$

y la **varianza**,

$$Var(X) = \frac{\beta^2\pi^2}{3}\,.$$

Ejemplo 10.11. Calculemos los deciles de una distribución logística estándar. El primer decil, d_1, valor de la variable que deja a su izquierda el 10 % de la probabilidad, se obtendrá igualando la función de distribución a 0.10; así,

$$F_Y(d_1) = 0.10\,,$$

o lo que es lo mismo,

$$\frac{1}{1+e^{-d_1}} = 0.10\,,$$

de donde resulta que

$$d_1 = -\ln(9) = -2.20\,.$$

Igualmente, obtendremos el i-ésimo decil ($i = 2, 3, \ldots, 9$) resolviendo la ecuación

$$F_Y(d_i) = \frac{i}{10}\,.$$

Entonces,

$$\begin{array}{llll} d_2 = -1.39 & d_3 = -0.85 & d_4 = -0.40 & d_5 = 0 \\ d_6 = 0.40 & d_7 = 0.85 & d_8 = 1.39 & d_9 = 2.20\,. \end{array}$$

Hallemos, ahora, los deciles de una distribución $N(0,1)$. Buscaremos en las tablas de la normal tipificada los valores d_i ($i = 1, \ldots, 9$) tales que

$$\Phi(d_i) = \frac{i}{10}\,,$$

resultando:

$$\begin{array}{llll} d_1 = -1.28 & d_2 = -0.84 & d_3 = -0.52 & d_4 = -0.25 \quad d_5 = 0 \\ d_6 = 0.25 & d_7 = 0.52 & d_8 = 0.84 & d_9 = 1.28\,. \end{array}$$

Obsérvese que la simetría de la distribución logística y de la normal hace que los deciles también presenten una estructura simétrica.

Como se puede comprobar comparando los deciles para ambas distribuciones, la distribución logística estándar no sólo está más dispersa globalmente que la normal estándar (la varianza de aquélla es $\frac{\pi^2}{3}$, mientras que la de ésta es 1), sino que esa dispersión es mejor a lo largo de todos los valores reales. ◁

10.10 Ejercicios propuestos

10.1. Sea X una variable aleatoria uniforme en el intervalo $(0,1)$. Compruébese que $E\left(\frac{1}{X}\right)$ no existe.

10.2. Sean X_1 y X_2 dos variables aleatorias independientes con distribución $U(0,1)$. ¿Cuál es la probabilidad de que la variable $\overline{X} = \frac{X_1+X_2}{2}$ esté comprendida entre $\frac{1}{4}$ y $\frac{3}{4}$?

10.3. Se eligen al azar e independientemente dos valores X e Y del intervalo $(0,1)$. Se definen las siguientes distancias: $U = |X - Y|$ y $V = (X - Y)^2$. Calcúlese la esperanza y varianza de U y V.

10.4. Se elige al azar un punto X en el intervalo $(0,2)$. Si el punto escogido es x, se escoge al azar otro punto Y en el intervalo $(0, x^2)$.

a) Sabiendo que el segundo punto elegido ha resultado ser el uno, calcúlese el valor esperado en la primera elección.

b) Hállese $p[X^2 < 2Y]$.

10.5. Sean X_1 y X_2 variables aleatorias independientes e igualmente distribuidas con distribución uniforme en el intervalo $(0,1)$. Para $i = 1, 2$, se definen las variables $Z_i = -2\ln(X_i)$. Determínese:

a) La distribución de Z_i.

b) La probabilidad del suceso $[Z_1 + Z_2 \leq 4]$.

10.6. El incremento anual de los salarios (en porcentaje) pactado en convenios colectivos para el período 1994-95 es una variable X con distribución $U(0,5)$. Todos los convenios se han firmado con la siguiente cláusula de revisión: si el IPC de 1994 supera el incremento salarial pactado, los salarios nominales subirán en 1995 lo necesario para que no haya una pérdida en el salario real. El gobierno prevé que la inflación para 1994 seguirá una distribución:

$$f_Y(y) = \begin{cases} \frac{-2y+10}{25} & \text{si } 0 \leq y \leq 5 \\ 0 & \text{en el resto} \end{cases}$$

Hállese el incremento esperado de los salarios nominales para 1995.

10.7. Sea Z una variable aleatoria con distribución $N(0,1)$. Demuéstrese que la variable $X = \sigma Z + \mu$, donde μ y σ son dos números reales cualesquiera ($\sigma > 0$), sigue una distribución $N(\mu, \sigma)$.

10.8. El número de horas que un estudiante necesita para aprender un tema de historia es una variable aleatoria con distribución $N(\mu, \sigma)$. Si el 84.13 % de los alumnos emplea más de tres horas y sólo el 2.28 % más de nueve, ¿cuánto valen μ y σ?

10.9. Sea X una variable aleatoria con distribución $N(0,1)$ y sea $Z = F_X(X)$. Demuéstrese que $E(Z) = \frac{1}{2}$.

10.10. El peso, en gramos, de los recién nacidos en un hospital está distribuido normalmente con media 3000 y desviación típica σ. ¿Qué valor tiene σ, si el 98 % de los bebés tiene un peso comprendido entre 2.5 y 3.5 kilos?

10.11. Sea X una variable aleatoria con distribución $N(0,\sigma)$. ¿Es $\sqrt{X^2}$ una variable normal?

10.12. La cotización en bolsa de un cierto título se considera una variable aleatoria normalmente distribuida con parámetros desconocidos, pero se dispone de la siguiente información:

- Existe un 2.5 % de probabilidad de que la cotización supere el índice 400.
- Hay un 50 % de probabilidad de que no llegue a superarse el índice 380.

Según los datos anteriores, ¿cuál es la probabilidad de que en un cierto momento la cotización se encuentre entre 385 y 395?

10.13. Sean $X_1,\ldots,X_n,Y_1,\ldots,Y_n$ variables aleatorias independientes, normales, de media μ y varianza 4. Sean $\overline{X}$ e $\overline{Y}$ las medias muestrales de las variables X_i e Y_i, respectivamente. ¿Qué valor debe tomar n para que $\overline{X}$ e $\overline{Y}$ difieran entre sí en menos de una unidad con una probabilidad de al menos 0.95?

10.14. En un proceso en serie se fabrican botellas de cola cuya capacidad responde a una distribución normal de media 1 litro y desviación típica 0.01. Se considera inutilizable toda botella con capacidad inferior a 0.99 litros o superior a 1.1 litros. ¿Qué porcentaje de botellas será rechazado?

10.15. Una compañía de seguros desea contratar dos nuevos agentes. Para ello realiza un proceso de selección sucesiva de los candidatos, consistente en un examen tipo test que, una vez superado, permitirá a los preseleccionados presentarse a una entrevista personal.
Para aprobar el test es necesario tener una puntuación mayor que 6. La puntuación de un individuo es una variable normal de media 4 y desviación típica 2. Además, se sabe que las mujeres pasan con éxito la entrevista el 80 % de las veces, siendo el porcentaje de 70 para el caso de los hombres.

a) ¿Cuál es la probabilidad de que sean seleccionados como agentes el segundo y tercer candidato?

b) Si ha sido necesario examinar a 50 candidatos antes de elegir al segundo agente, ¿cuál es la probabilidad de que se hayan realizado 20 entrevistas antes de elegir al primero?

10.16. Sean $X_1,\ldots,X_n$ variables aleatorias independientes con distribución $N(2,1)$. Hállese el valor de n para que se cumpla:

$$p\left[\sum_{i=1}^{n} X_i \leq 2\right] = 0.95\ .$$

10.17. Sea X una variable aleatoria $N(0,1)$. Obténgase la distribución de la variable aleatoria $Y = e^X$.

10.18. Los ingresos mensuales, en miles de pesetas, de los individuos de un país siguen una distribución logarítmico-normal de parámetros $\mu = 3$ y $\sigma = 2$. ¿Qué porcentaje tiene una renta superior a la media?

10.19. Sea X una variable aleatoria con distribución logarítmico-normal de parámetros μ y σ. Obténgase la moda y la mediana y compárense las tres medidas de posición central.

10.20. Estúdiese el crecimiento y decrecimiento de la función de densidad de una distribución gamma y obténgase su moda.

10.21. El tiempo, en horas, empleado diariamente en transporte por los trabajadores de una gran ciudad, sigue una distribución gamma con parámetros $a = 2$ y $p = 2$.

a) Calcúlese la probabilidad de que un trabajador emplee más de media hora en transporte.

b) Los trabajadores de un barrio periférico emplean al menos una hora, ¿cuál es la probabilidad de que no superen la hora y media?

c) Determínese el tiempo mínimo que emplea el 50 % de los trabajadores que más tiempo pierden en transporte.

10.22. Una compañía aseguradora ofrece a sus clientes una póliza de accidente con las siguientes características:

- La compañía abona la totalidad del siniestro hasta una cuantía de 3 mil pesetas.
- Si la cuantía está comprendida entre 3 y 70 mil pesetas, abona íntegramente las 3 mil primeras y la mitad de lo que exceda.
- Si la cantidad supera las 70 mil pesetas, paga 36.5 mil pesetas.

Se ha estimado que la cuantía de daños por accidente sigue una distribución gamma de parámetros $a = 0.10$ y $p = 2$. Hállese la cantidad que la compañía espera pagar por accidente.

10.23. Sea X una variable aleatoria con distribución exponencial. Calcúlese $p[X > E(X)]$ y obténgase la mediana y el coeficiente de variación.

10.24. Sea X una variable aleatoria con distribución exponencial negativa.

a) Calcúlese $p[|X - \mu| \geq k \cdot \sigma]$ para $k > 0$.

b) Compárese le resultado anterior con el que obtendríamos utilizando la desigualdad de Chebyshev.

10.25. Una fábrica utiliza dos métodos para la producción de bombillas: el 70 % de ellas las fabrica por el método A y el resto por el método B. La duración de las bombillas sigue una distribución exponencial de media 80 o 100 horas, según se utilice el método A o el B.
Se toman 10 bombillas al azar, ¿cuál es la probabilidad de que 6 de ellas duren al menos 90 horas?

10.26. El coste de una cierta pieza es de 25 pesetas. El precio de venta depende del diámetro interior de la pieza, X, variable exponencial negativa de media λ. Si el diámetro es mayor que 3 o menor que 1, se desecha la pieza, y si el diámetro está comprendido entre estos límites, se vende al precio de 40 pesetas. La máquina que produce la pieza tiene un dispositivo que permite ajustar el diámetro medio de las piezas, λ. Hállese el valor de λ que maximice el beneficio medio.

10.27. En un laboratorio, el 40 % de los ratones sometidos a un cierto estímulo son machos. El tiempo, en minutos, que un ratón tarda en responder al estímulo es una variable aleatoria exponencial, siendo el tiempo medio de respuesta el doble en los machos que en las hembras.
Se toma un ratón al azar y se le somete al experimento. Si la probabilidad de que tarde más de 4 minutos en responder es 0.8, obténganse los tiempos medios de respuesta para los machos y hembras.

10.28. El tiempo de vida de una válvula es una variable exponencial negativa de media λ. Elegidas dos válvulas al azar, y suponiendo que operan independientemente, se sabe que la probabilidad de que las dos duren al menos 4 horas es de 0.252. Calcúlese $E(U)$, siendo U la variable que representa el tiempo de vida de la primera pieza que falla en el par.

10.29. La Compañía Telefónica cobra 80 pesetas por cada minuto o fracción de llamada interurbana. Hállese el ingreso esperado por llamada, si se ha evaluado que la duración de una llamada, en minutos, sigue una distribución exponencial negativa de parámetro uno.

10.30. Dada una variable aleatoria con distribución $U(0,1)$, demuéstrese que la variable aleatoria $Y = \ln\left(\frac{1}{X}\right)$ sigue una distribución $\mathcal{E}(1)$.

10.31. La variable aleatoria X_t, número de clientes que acuden a un establecimiento en un intervalo de tiempo $(0,t)$, sigue una distribución de Poisson de parámetro λt. Sea T el tiempo de espera hasta la llegada del primer cliente. Hállese:

a) $p[T > t]$.

b) ¿Cuál será el tiempo esperado hasta la llegada del primer cliente?

10.32. Sea X una variable aleatoria con distribución beta $\beta(2,2)$. Obténgase la media, la moda, la mediana y el coeficiente de variación.

10.33. La proporción de alumnos de una facultad que aprueban la asignatura de Estadística cada año, es una variable aleatoria con distribución beta de parámetros $p = 3$ y $q = 2$.

a) Calcúlese la probabilidad de que, en un año cualquiera, suspendan menos del 20 %.

b) Hállese la probabilidad de que la proporción de estudiantes que aprueban en un año sea menor que la proporción media de aprobados.

10.34. Una variable aleatoria X tiene como función de densidad:

$$f_X(x) = \begin{cases} \frac{5}{3}\left(\frac{3}{x}\right)^6 & x \geq x_0 \\ 0 & \text{en el resto} \end{cases}$$

a) Determínese el valor de x_0.
b) ¿Qué distribución sigue X?
c) Calcúlese $E(X)$.

10.35. En una distribución de Pareto con $\alpha \geq 3$, si x_0 se mantiene constante y crece α, ¿qué ocurre con la esperanza y la varianza? Realícese el mismo análisis cuando α se mantiene constante y crece x_0.

10.36. Sea X una variable aleatoria con distribución de Pareto de parámetros α y x_0.

a) Obténgase el coeficiente de variación y la mediana.
b) Calcúlese $p[X < E(X)]$.

10.37. Sea Y una variable aleatoria con distribución de Cauchy estándar.

a) Demuéstrese que la variable $X = \beta \cdot Y + \alpha$, donde α y β son dos números reales cualesquiera ($\beta > 0$), sigue una distribución de Cauchy de parámetros α y β.
b) Obténgase la función de distribución de una variable aleatoria con distribución de Cauchy de parámetros α y β.

10.38. Hállese la moda y la mediana de una distribución de Cauchy.

10.39. Sea X una variable aleatoria con distribución de Cauchy. Se considera la variable aleatoria $Y = \frac{1}{X}$; obténgase $E(Y)$.

10.40. Sean X e Y dos variables aleatorias independientes con distribución $N(0,1)$. Demuéstrese que $Z = \frac{X}{Y}$ sigue una distribución de Cauchy.

10.41. Demuéstrese que la función de densidad de una distribución logística es simétrica respecto de α. Obténgase la moda y la mediana.

10.42. Hallar el intervalo intercuartílico de una distribución logística estándar y compárese con la de una distribución de Cauchy estándar.

Capítulo 11

DISTRIBUCIONES RELACIONADAS CON LA NORMAL

En este capítulo estudiaremos varias distribuciones continuas, construidas a partir de la distribución normal que, aunque serán usadas como distribuciones tipo (especialmente la distribución χ^2 de Pearson), no fueron diseñadas con este objetivo, sino para asignar medidas probabilísticas a las variables aleatorias que aparecen en los procedimientos de estimación.

Uno de los problemas fundamentales de la Inferencia Estadística es la estimación de los parámetros que contienen las distribuciones tipo. Esta estimación la realizaremos a partir de los estadísticos, que son funciones de varias variables aleatorias independientes e igualmente distribuidas. Bajo hipótesis de normalidad, que es una hipótesis frecuentemente utilizada en este contexto, los principales estadísticos siguen distribuciones de las estudiadas en este capítulo.

Como observará el lector, la descripción de las variables se desarrolla con distintas pautas a las utilizadas en los dos capítulos anteriores. Para las tres distribuciones estudiadas, interesa más conocer su relación con la distribución normal que su función de densidad y momentos. En otras distribuciones tipo, la función de densidad tiene un interés fundamental, tanto en el cálculo de probabilidades como por el hecho de que el conocimiento de su perfil sirve de guía en la modelización de fenómenos aleatorios. Sin embargo, las distribuciones relacionadas con la normal tienen escasa utilidad en la modelización y, además, sus probabilidades se calculan mediante las tablas que presentamos en el anexo final del libro.

11.1 Distribución χ^2 de Pearson

Esta distribución surge cuando se plantea estudiar la suma de los cuadrados de variables independientes e igualmente distribuidas con distribución normal.

Sea X una variable aleatoria con distribución $N(0,1)$; vamos a hallar la distribución de $Y = X^2$. Calculemos, en primer lugar, la función de distribución.

Obviamente, $F_Y(y)$ vale cero, si y es menor o igual que cero, ya que la variable sólo toma valores no negativos. Si $y \geq 0$,

$$\begin{aligned} F_Y(y) = p[Y \leq y] &= p[X^2 \leq y] = p[-\sqrt{y} \leq X \leq \sqrt{y}] = \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \,. \end{aligned}$$

Ahora bien, por la simetría de la distribución normal,

$$F_Y(y) = F_X(\sqrt{y}) - [1 - F_X(\sqrt{y})] = 2 \cdot F_X(\sqrt{y}) - 1 \,.$$

Derivando la función de distribución obtenemos la función de densidad de Y, que valdrá cero, si y es negativo, y

$$\begin{aligned} f_Y(y) &= f_X(\sqrt{y}) \cdot \frac{1}{\sqrt{y}} = \frac{1/\sqrt{2}}{\sqrt{\pi}} \cdot y^{-1/2} \cdot e^{-y/2} = \\ &= \frac{(1/2)^{1/2}}{\Gamma(1/2)} \cdot y^{\frac{1}{2}-1} \cdot e^{-y/2} \,, \end{aligned}$$

si $y > 0$, sin más que tener en cuenta que $\Gamma(1/2) = \sqrt{\pi}$.

En definitiva, la función de densidad de la variable $Y = X^2$ vale

$$f_Y(y) = \begin{cases} \frac{(1/2)^{1/2}}{\Gamma(1/2)} \cdot y^{\frac{1}{2}-1} \cdot e^{-y/2} & y > 0 \\ 0 & \text{en el resto} \end{cases}$$

Diremos que Y sigue una **distribución χ^2 (ji cuadrado) con un grado de libertad** (más adelante explicaremos los motivos de esta denominación). Lo indicaremos abreviadamente como

$$Y \hookrightarrow \chi^2_1 \,.$$

Como puede comprobar el lector comparando las funciones de densidad, la distribución χ^2_1 no es sino una gamma, cuyos dos parámetros son iguales a $\frac{1}{2}$. Este hecho, que denotaremos como

$$\chi^2_1 \equiv \gamma\left(\frac{1}{2}, \frac{1}{2}\right) \,,$$

ayudará a obtener sus principales características.

La definición anterior se generaliza a la suma de los cuadrados de cualquier cantidad de variables normales estándar que sean independientes.

Definición 11.1 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes e igualmente distribuidas con distribución $N(0,1)$. Entonces, decimos que la variable $X = X_1^2 + \cdots + X_n^2$ sigue una distribución χ^2 con n grados de libertad.*

Si una variable sigue esta distribución lo indicaremos abreviadamente como

$$X \hookrightarrow \chi^2_n \,.$$

La distribución χ^2_n tiene un único parámetro, que son sus grados de libertad. Entenderemos el porqué de esta denominación cuando utilicemos esta distribución en Inferencia Estadística.

Como las variables $X_1, \ldots, X_n$ que aparecen en la definición son independientes con distribución $N(0,1)$, sus cuadrados, $X_1^2, \ldots, X_n^2$, son también independientes, y la distribución de cada uno de ellos es una χ^2_1. Por tanto, una χ^2 con n grados de libertad puede interpretarse como la suma de n variables χ^2 con un grado de libertad e independientes.

Dado que las distribuciones χ^2_1 son, realmente, distribuciones $\gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ y que, como vimos en la proposición 10.7, la distribución gamma es reproductiva, una variable χ^2_n, suma de n variables $\gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ e independientes, tiene una distribución $\gamma\left(\frac{1}{2}, \frac{n}{2}\right)$. Es decir,

$$\chi^2_n \equiv \gamma\left(\frac{1}{2}, \frac{n}{2}\right) .$$

Aunque desde el punto de vista operativo carece de interés, damos a continuación la expresión de su **función de densidad**:

$$f_X(x) = \begin{cases} \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{\frac{-1}{2}x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

En la figura 11.1 se representa la función de densidad para distintas distribuciones χ^2.

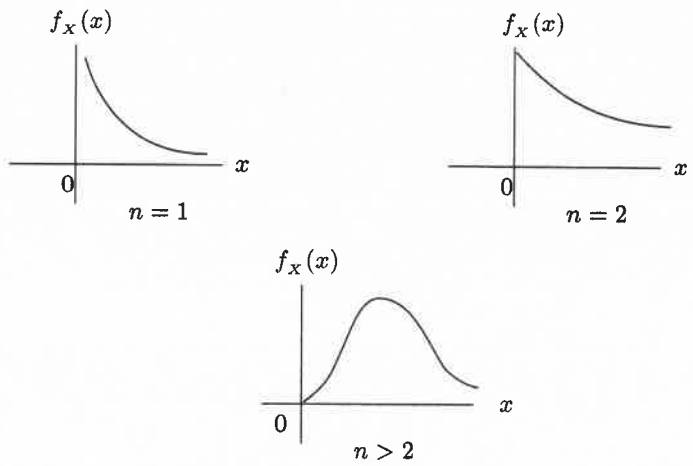


FIG. 11.1.

En el anexo final del libro se recoge una tabla con los valores de la **función de distribución** de una χ^2 para distintos grados de libertad, que permitirá hallar sus probabilidades. Para valores grandes de n utilizaremos una aproximación a la

distribución normal; en concreto, si X es una χ^2_n, la variable $Y = \sqrt{2 \cdot X}$ sigue aproximadamente una distribución normal $N(\sqrt{2n-1}, 1)$.

Hallemos los momentos y la función característica de esta variable, a partir de los resultados obtenidos para las distribución gamma; así, su **esperanza** vale

$$E(X) = \frac{n/2}{1/2} = n \,,$$

es decir, los grados de libertad. Su **varianza**,

$$Var(X) = \frac{n/2}{(1/2)^2} = 2n \,,$$

es el doble de los grados de libertad, y la **función característica** es

$$\varphi_X(t) = (1 - 2it)^{\frac{-n}{2}} \,.$$

La reproductividad de las variables gamma garantiza la de las χ^2, como muestra la siguiente proposición.

Proposición 11.1 *La suma de variables χ^2 independientes sigue una distribución χ^2 cuyos grados de libertad son la suma de los grados de libertad de los sumandos.*

Demostración. Sean $X_1, \ldots, X_m$, m variables independientes tales que $X_i \hookrightarrow \chi^2_{n_i}$ para $i = 1, \ldots, m$. Como $\chi^2_{n_i} \equiv \gamma\left(\frac{1}{2}, \frac{n_i}{2}\right)$, y las gamma son reproductivas cuando el primer parámetro permanece constante, resulta que su suma es también una gamma de parámetros $\frac{1}{2}$ y $\frac{n_1+\cdots+n_m}{2}$, esto es, una χ^2 con $n_1 + \cdots + n_m$ grados de libertad. ♣

La siguiente proposición es, en cierto modo, recíproca de la anterior. Si la distribución de la suma de dos variables es una χ^2, y una de ellas lo es, o bien ésta coincide con la suma (el segundo sumando sería igual a cero), o bien la segunda variable es también una χ^2.

Proposición 11.2 *Sean dos variables independientes, X_1 y X_2, tales que $X_1 \hookrightarrow \chi^2_{n_1}$ y $X_1 + X_2 \hookrightarrow \chi^2_n$, con $n > n_1$. Entonces,*

$$X_2 \hookrightarrow \chi^2_{n-n_1} \,.$$

Demostración. En virtud de la independencia entre X_1 y X_2, la función característica de la suma es el producto de las funciones características de los sumandos,

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t) \,,$$

con lo cual,

$$\varphi_{X_2}(t) = \frac{\varphi_{X_1+X_2}(t)}{\varphi_{X_1}(t)} = \frac{(1-2it)^{-n/2}}{(1-2it)^{-n_1/2}} = (1-2it)^{-(n-n_1)/2} \,.$$

Como $n > n_1$, su diferencia, $n - n_1$, es un entero positivo y, por tanto, esta función característica es la de una $\chi^2_{n-n_1}$, lo que, con la ayuda del teorema de unicidad (corolario 8.6.1), prueba el resultado. ♣

Observe el lector que, si $n = n_1$, $\varphi_{X_2}(t) = 1$, para todo número real, t, y en consecuencia, X_2 es una variable degenerada en cero. Además, se demuestra (no lo haremos en este libro) que n nunca puede ser menor que n_1, ya que, entonces, φ_{X_2} no sería una función característica[1].

El resultado siguiente permite obtener probabilidades para algunas variables gamma, utilizando las tablas de la distribución χ^2.

Proposición 11.3 *Sea* $X \hookrightarrow \gamma(a,p)$, *siendo* $2p$ *un número entero positivo. Entonces,*

$$2aX \hookrightarrow \chi^2_{2p} .$$

La demostración es inmediata a partir de la proposición 10.8.

Ejemplo 11.1. El tiempo que un funcionario emplea en atender a una persona es una variable aleatoria con distribución exponencial negativa de media 4 minutos. Delante de un individuo hay una cola formada por 10 personas. ¿Cuál es la probabilidad de que tenga que esperar más de 31 minutos hasta que le toque el turno?

Consideremos las variables X_i, $i = 1, \ldots, 10$, tiempo que se emplea en atender a cada uno de los individuos que están por delante en la cola. Por tanto, la variable

$$Y = X_1 + \cdots + X_{10}$$

representará el tiempo que le falta al individuo para ser atendido. La probabilidad que tenemos que hallar es

$$p[Y > 31] .$$

Veamos cuál es la distribución de Y. Dado que X_i sigue una distribución exponencial negativa de media 4, también la podemos considerar como una gamma. Así,

$$X_i \hookrightarrow \mathcal{E}\left(\frac{1}{4}\right) \equiv \gamma\left(\frac{1}{4}, 1\right) .$$

Entonces, teniendo en cuenta la propiedad reproductiva de las distribuciones gamma,

$$Y = X_1 + \cdots + X_{10} \hookrightarrow \gamma\left(\frac{1}{4}, 10\right) .$$

Ahora bien, la función de densidad de esta gamma no va a servir para obtener la probabilidad pedida, pues no encontraremos su primitiva. Tenemos una alternativa: buscar el valor de la probabilidad en las tablas correspondientes o convertir la gamma en una χ^2, si es posible, y buscar en sus tablas, siempre más asequibles, el valor de la probabilidad. Elijamos este camino.

Utilizando la proposición anterior, la distribución de $\frac{1}{2}Y$ será

$$\frac{1}{2}Y \hookrightarrow \gamma\left(\frac{1}{2}, 10\right) \equiv \gamma\left(\frac{1}{2}, \frac{20}{2}\right) \equiv \chi^2_{20} .$$

Por tanto,

$$p[Y > 31] = p\left[\frac{Y}{2} > 15.5\right] = 1 - p\left[\frac{Y}{2} \leq 15.5\right] .$$

1. Toda función característica ha de cumplir que $|\varphi_X(t)| \leq 1$ (ejercicio propuesto 8.1). Si n fuese menor que n_1, la función característica φ_{X_2} resultante no verificaría la propiedad anterior.

Las tablas de la χ^2 proporcionarán la probabilidad de la función de distribución de una χ^2 con 20 grados de libertad en el punto 15.5; esta probabilidad[2] es 0.25.

Entonces, la probabilidad de que la persona tenga que esperar más de 31 minutos es 0.25. ◁

Cuando las variables normales que sirven para construir la ditribución χ^2 no tienen esperanza nula, la distribución se denomina χ^2 descentrada[3]. Dicha distribución se emplea para calcular la potencia de ciertos contrastes de hipótesis, un concepto que definiremos en el volumen dedicado a la Inferencia Estadística.

Definición 11.2 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes e igualmente distribuidas con distribución $N(\mu_i, 1)$. La variable $X_1^2 + \cdots + X_n^2$ se dice que sigue una distribución χ^2 descentrada, con n grados de libertad y parámetro de descentralización $\delta^2 = \sum_{i=1}^n \mu_i^2$. La denotaremos abreviadamente como $\chi_n^2(\delta^2)$.*

11.2 Distribución t de Student

Definición 11.3 *Sean X y $X_1, \ldots, X_n$, $n+1$ variables aleatorias independientes e igualmente distribuidas con distribución $N(0,1)$. Decimos, entonces, que la variable*

$$T = \frac{X}{\sqrt{\frac{X_1^2 + \cdots + X_n^2}{n}}}$$

sigue una distribución t de Student con n grados de libertad.

Si una variable, T, sigue una distribución t de Student con n grados de libertad, lo indicaremos de forma abreviada como

$$T \hookrightarrow t_n \, .$$

Como $X_1^2 + \cdots + X_n^2$ sigue una distribución χ^2 con n grados de libertad, puede darse una definición equivalente de esta distribución de probabilidad.

Definición 11.4 *Sea X una variable aleatoria con distribución $N(0,1)$ y sea Y una variable aleatoria con distribución χ_n^2. Si X e Y son independientes, la variable*

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

se dice que sigue una distribución t de Student con n grados de libertad.

2. El lector advertirá al consultar las tablas que éstas están diseñadas para hallar los valores de la variable que, en una χ^2 con n grados de libertad, acumulan una cantidad p de probabilidad, resultando *incómodas* para el proceso contrario. La justificación de este hecho está en el uso de la distribución χ^2 como distribución en la Inferencia Estadística con variables normales.

3. No la estudiaremos en este libro, y el lector interesado en ella puede acudir, por ejemplo, a N. L. JOHNSON y S. KOTZ, *Continuous Univariate Distributions-2*, Houghton Mifflin Co., Boston, 1970, pp. 130-146.

La distribución t de Student tiene un único parámetro, que indica los grados de libertad de la χ^2 que aparece en el denominador.

El origen de esta distribución se encuentra en la estimación de esperanzas de distribuciones normales, cuando su desviación típica es desconocida. W. S. GOSSET, por seudónimo *Student*, un industrial cervecero, la propuso y tabuló en 1908.

Seguidamente hallamos la expresión de la **función de densidad** de esta variable; como antes, dicha función tiene poco interés práctico.

Proposición 11.4 *La función de densidad de una variable aleatoria con distribución t de Student con n grados de libertad es*

$$f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \cdot \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \quad \textit{para} \ -\infty < t < \infty \, .$$

Demostración. Sea X una variable con distribución $N(0,1)$ y sea $Y = \sqrt{\frac{U}{n}}$, donde U es una variable χ^2_n, y tal que X y U son independientes. Calculemos, en primer lugar, la función de distribución de Y,

$$F_Y(y) = p\left[\sqrt{\frac{U}{n}} \leq y\right] = p[U \leq ny^2] = F_U(ny^2) \, .$$

Derivando la expresión anterior, y sustituyendo el valor de la función de densidad de una χ^2, obtenemos:

$$f_Y(y) = f_U(ny^2) \cdot 2ny = \begin{cases} \dfrac{2(n/2)^{n/2}}{\Gamma(n/2)} \cdot y^{n-1} \cdot e^{-\frac{ny^2}{2}} & y > 0 \\ 0 & \text{en el resto} \end{cases}$$

Como X e Y son independientes, su función de densidad conjunta es el producto de las marginales; así,

$$f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y) = \frac{2(n/2)^{n/2}}{\sqrt{2\pi} \cdot \Gamma(n/2)} \cdot y^{n-1} \cdot e^{-\frac{x^2+ny^2}{2}} \, ,$$

para $-\infty < x < \infty$ e $y > 0$.

Con objeto de obtener la función de densidad del cociente entre X e Y, hacemos el siguiente cambio de variable:

$$T = \frac{X}{Y}$$

$$Z = Y \, .$$

El módulo del determinante jacobiano de esta transformación es $|J| = z$, con lo que, en virtud del teorema del cambio de variable, resulta

$$f_{(T,Z)}(t,z) = \frac{2(n/2)^{n/2}}{\sqrt{2\pi} \cdot \Gamma(n/2)} \cdot z^n \cdot e^{-\frac{z^2(t^2+n)}{2}} \, ,$$

donde $-\infty < t < \infty$ y $z > 0$. A partir de esta función de densidad conjunta, integrando, obtenemos[4] la función de densidad marginal de T,

$$f_T(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \cdot \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} \quad \text{para } -\infty < t < \infty \,. \clubsuit$$

Esta función de densidad tiene un perfil similar al de la función de densidad de la distribución $N(0,1)$. Es simétrica con respecto al origen de coordenadas $(f_T(t) = f_T(-t))$, y para ese valor, $t = 0$, alcanza su máximo. Por tanto, la mediana y la moda de la distribución valen cero. Su aspecto es, asimismo, acampanado, con dos puntos de inflexión simétricos con respecto del origen. La figura 11.2 muestra su representación gráfica.

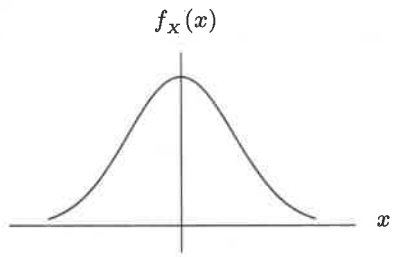


FIG. 11.2.

De hecho, su función de densidad es algo más aplastada cerca del origen que la de la $N(0,1)$, y toma valores mayores que ésta, cuando se aleja del origen de coordenadas. Según aumentan los grados de libertad, su perfil se hace más y más parecido y, como indicaremos más adelante, para $n > 30$ son prácticamente indistinguibles.

Las mayores diferencias se presentan cuando el número de grados de libertad es pequeño. En concreto, para $n = 1$ la distribución t_1 coincide con la de Cauchy estándar, como indica la siguiente proposición.

Proposición 11.5 *La distribución de la t de Student con 1 grado de libertad coincide con la distribución de Cauchy estándar.*

Es evidente, sin más que escribir la función de densidad $f_T(t)$ obtenida en la proposición anterior para $n = 1$.

4. Para resolver la integral correspondiente, basta hacer el cambio de variable $z^2 = w$ y, multiplicando por las constantes adecuadas, tratar de que aparezca como integrando la función de densidad de una distribución gamma.

La **función de distribución** de una distribución t de Student,

$$F_T(t) = \int_{-\infty}^{t} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \cdot \frac{1}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}} dx \qquad \text{para } -\infty < t < \infty ,$$

está tabulada; en el anexo final se recoge dicha tabla.

Como hemos sugerido en un comentario anterior, cuando n tiende a infinito, existe una aproximación de la distribución t de *Student* a la normal estándar. En la práctica, a partir de 30 grados de libertad sus funciones de distribución pueden considerarse casi idénticas, pudiéndose utilizar las tablas de la normal para calcular probabilidades de la t de Student.

Indiquemos los principales momentos de la distribución. Evidentemente, de acuerdo con el corolario anterior, para $n = 1$ no existe la esperanza. Para $n > 1$ se demuestra que la **esperanza** existe, y vale

$$E(X) = 0 ,$$

debido a la simetría de la función de densidad.

Asimismo, para $n = 1$ o $n = 2$ no existe la varianza. Si $n > 2$, la **varianza** es

$$Var(X) = \frac{n}{n-2} .$$

Al igual que en el caso de la distribución χ^2, puede definirse la distribución t descentrada. Para esta distribución son válidos los comentarios acerca de su utilidad que hicimos al definir la χ^2 descentrada. Demos únicamente su definición[5].

Definición 11.5 *Sea X es una variable aleatoria con distribución $N(\mu, 1)$, e Y una variable aleatoria con distribución χ^2_n. Si X e Y son independientes, decimos que la variable*

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

sigue una distribución t de Student descentrada, con n grados de libertad. El valor μ se denomina parámetro de descentralización.

Si T sigue una distribución t descentrada, con n grados de libertad y parámetro de descentralización μ, lo denotamos abreviadamente por

$$T \hookrightarrow t_n(\mu) .$$

5. El lector interesado en su estudio puede encontrar un desarrollo completo de esta distribución en N. L. JOHNSON y S. KOTZ, *Continuous Univariate Distributions-2*, Houghton Mifflin Co., Boston, 1970, pp. 201 a 220.

11.3 Distribución F de Snedecor

Definición 11.6 *Sean $X_1, \ldots, X_{n_1}$ e $Y_1, \ldots, Y_{n_2}$, $n_1 + n_2$ variables aleatorias independientes con distribución $N(0,1)$. Decimos que la variable*

$$U = \frac{\frac{X_1^2 + \cdots + X_{n_1}^2}{n_1}}{\frac{Y_1^2 + \cdots + Y_{n_2}^2}{n_2}}$$

sigue una distribución F de Snedecor con n_1 y n_2 grados de libertad.

Como las sumas de cuadrados que aparecen en el numerador y en el denominador de la expresión anterior siguen distribuciones χ^2, podemos dar una definición equivalente.

Definición 11.7 *Sean X e Y variables aleatorias independientes con distribuciones χ^2 de n_1 y n_2 grados de libertad, respectivamente. La variable*

$$U = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}}$$

se dice que sigue una distribución F de Snedecor con n_1 y n_2 grados de libertad.

Si una variable X sigue una distribución F con n_1 y n_2 grados de libertad, lo denotaremos de forma abreviada como

$$X \hookrightarrow F_{n_1,n_2} \, .$$

La distribución F de Snedecor tiene dos parámetros, que son los grados de libertad de las χ^2. Se dice que n_1 son los grados de libertad del numerador y n_2 los del denominador.

Esta distribución aparece en relación con las técnicas de análisis de la varianza, que se dedican a contrastar la igualdad del comportamiento de una variable aleatoria dentro de varias poblaciones.

Obtengamos la función de densidad de una distribución F de Snedecor, aunque, como en las distribuciones anteriores, no sea operativa.

Proposición 11.6 *La función de densidad de una variable aleatoria con distribución F de Snedecor con n_1 y n_2 grados de libertad es*

$$f_U(u) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(n_1/2) \cdot \Gamma(n_2/2)} \cdot n_1^{n_1/2} \cdot n_2^{n_2/2} \cdot u^{\frac{n_1}{2}-1} \cdot (n_2 + n_1 u)^{-\frac{n_1+n_2}{2}}, \quad \textit{para } u > 0 \, ,$$

anulándose en el resto.

Demostración: Sean $T \hookrightarrow \chi^2_{n_1}$ y $Z \hookrightarrow \chi^2_{n_2}$ variables independientes. Dadas $X = \frac{T}{n_1}$ e $Y = \frac{Z}{n_2}$, se trata de hallar la función de densidad de $U = \frac{X}{Y}$.

De las propiedades demostradas para las distribuciones χ^2 y gamma se deduce que

$$X \hookrightarrow \gamma\left(\frac{n_1}{2}, \frac{n_1}{2}\right)$$

e

$$Y \hookrightarrow \gamma\left(\frac{n_2}{2}, \frac{n_2}{2}\right) .$$

Al ser X e Y independientes,

$$f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y) = \frac{(n_1/2)^{n_1/2} \cdot (n_2/2)^{n_2/2}}{\Gamma(n_1/2) \cdot \Gamma(n_2/2)} \cdot x^{\frac{n_1}{2}-1} \cdot y^{\frac{n_2}{2}-1} \cdot e^{-\frac{n_1 x + n_2 y}{2}} ,$$

para $x > 0$ e $y > 0$. Hagamos el cambio de variable

$$U = \frac{X}{Y}$$

$$V = Y .$$

El módulo del determinante jacobiano de esta transformación es $|J| = v$, con lo que en virtud del teorema del cambio de variable, obtenemos:

$$f_{U,V}(u,v) = \frac{(n_1/2)^{n_1/2} \cdot (n_2/2)^{n_2/2}}{\Gamma(n_1/2) \cdot \Gamma(n_2/2)} \cdot u^{\frac{n_1}{2}-1} \cdot v^{\frac{n_1+n_2}{2}-1} \cdot e^{-\frac{(n_1 u + n_2)v}{2}} ,$$

para $u > 0, v > 0$. Integrando respecto a v, obtenemos la marginal buscada. ♣

La figura 11.3 representa la función de densidad de una variable aleatoria con distribución F de Snedecor.

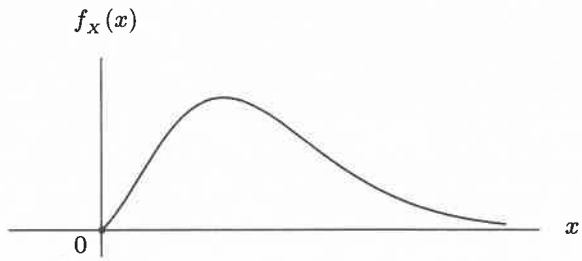


FIG. 11.3.

La **función de distribución** de una F de Snedecor está tabulada. Las tablas que aparecen en el anexo final de este libro proporcionan los valores de la variable F de Snedecor para los cuales la función de distribución vale 0.90, 0.95 y 0.99, para distintos grados de libertad del numerador y del denominador. En general, estas tablas son suficientes para cubrir las necesidades que se plantean en las aplicaciones prácticas. Si el lector necesita calcular otras probabilidades para una variable F de Snedecor, debe utilizar la siguiente proposición.

Proposición 11.7 *Sea X una variable aleatoria con distribución F de Snedecor de n_1 y n_2 grados de libertad. Entonces, la variable aleatoria*

$$Y = \frac{n_1 X}{n_2 + n_1 X}$$

sigue una distribución beta, con parámetros $\frac{n_1}{2}$ y $\frac{n_2}{2}$.

Demostración. Sea X una variable aleatoria con distribución F_{n_1,n_2}. La transformación

$$Y = \frac{n_1 X}{n_2 + n_1 X}$$

es estrictamente creciente (puede comprobar el lector que la derivada de $y = \frac{n_1 x}{n_2 + n_1 x}$ es estrictamente positiva), siendo su transformación inversa

$$X = \frac{n_2 Y}{n_1(1-Y)} .$$

Por tanto, para un valor y entre cero y uno,

$$F_Y(y) = p[Y \leq y] = p\left[\frac{n_1 X}{n_2 + n_1 X} \leq y\right] = p\left[X \leq \frac{n_2 y}{n_1(1-y)}\right] = F_X\left(\frac{n_2 y}{n_1(1-y)}\right) .$$

Derivando, hallamos la función de densidad:

$$f_Y(y) = f_X\left(\frac{n_2 y}{n_1(1-y)}\right) \cdot \frac{n_2}{n_1(1-y)^2} .$$

Sustituyendo por la función de densidad de X, obtenemos:

$$f_Y(y) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(n_1/2)\cdot\Gamma(n_2/2)} \cdot y^{\frac{n_1}{2}-1} \cdot (1-y)^{\frac{n_2}{2}-1} , \qquad \text{para} \quad 0 < y < 1 .$$

Esta función de densidad es la de la distribución beta $\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$. ♣

Como acabamos de decir, esta proposición sirve para utilizar las tablas de la distribución beta, cuya referencia dimos en la sección 10.6, a la hora de hallar probabilidades para una distribución F. Puesto que la función

$$y = \frac{n_1 x}{n_2 + n_1 x}$$

es una función estrictamente creciente, si X sigue una distribución F_{n_1,n_2} y x es positivo,

$$p[X \leq x] = p\left[\frac{n_1 X}{n_2 + n_1 X} \leq \frac{n_1 x}{n_2 + n_1 x}\right] = p\left[Y \leq \frac{n_1 x}{n_2 + n_1 x}\right] ,$$

donde Y es la variable descrita en la proposición anterior, esto es, una $\beta(\frac{n_1}{2}, \frac{n_2}{2})$.

El lector puede intuir el resultado recíproco cuya demostración es inmediata.

Proposición 11.8 *Sea X una variable aleatoria con distribución beta de parámetros p y q, siendo $2p$ y $2q$ números enteros. La variable*

$$Y = \frac{qX}{p(1-X)}$$

sigue una distribución F de Snedecor con $2p$ y $2q$ grados de libertad.

Así como la proposición 11.7 permite hallar probabilidades de una distribución F de Snedecor mediante las tablas de la distribución beta, la proposición 11.8 permite obtener las probabilidades de ciertas betas con las tablas de la F de Snedecor.

A continuación daremos los valores de los principales momentos, sin calcularlos. Para $n_1 = 1$ o $n_1 = 2$, no existe la esperanza. Para $n_2 > 2$, se demuestra que la **esperanza** vale

$$E(X) = \frac{n_2}{n_2 - 2} .$$

La **varianza**, que no existe si $n_2 = 1, 2, 3, 4$, para $n_2 > 4$ vale

$$Var(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 4)(n_2 - 2)^2} .$$

Existe una relación entre las variables con distribución F_{n_1,n_2} y F_{n_2,n_1}. De hecho, una es la inversa de la otra. Este resultado permite relacionar sus probabilidades.

Proposición 11.9 *Si $X \hookrightarrow F_{n_1,n_2}$ e $Y \hookrightarrow F_{n_2,n_1}$, entonces,*

$$p[X \leq x] = 1 - p\left[Y \leq \frac{1}{x}\right] \qquad \text{para } x > 0 .$$

Demostración. De acuerdo con la definición 11.6, si $X \hookrightarrow F_{n_1,n_2}$, entonces su inversa, $Y = \frac{1}{X} \hookrightarrow F_{n_2,n_1}$. Por tanto,

$$p[X \leq x] = p\left[\frac{1}{X} \geq \frac{1}{x}\right] = p\left[Y \geq \frac{1}{x}\right] = 1 - p\left[Y \leq \frac{1}{x}\right] ,$$

de donde se deduce el resultado. ♣

Puede definirse una distribución F de Snedecor descentrada, que resulta de utilizar en su numerador una χ^2 descentrada. Daremos sólo la definición[6].

Definición 11.8 *Sea X una variable aleatoria con distribución χ^2 descentrada con n_1 grados de libertad y parámetro de descentralización δ. Sea Y una variable*

6. El lector interesado puede encontrar un desarrollo completo de esta distribución en N. L. JOHNSON y S. KOTZ, *Continuous Univariate Distributions-2*, Houghton Mifflin Co., Boston, 1970, pp. 189–219.

aleatoria con distribución χ^2 con n_2 grados de libertad e independiente de X. La variable

$$V = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}}$$

se dice que sigue una distribución F de Snedecor descentrada con n_1 y n_2 grados de libertad y parámetro de descentralización δ.

Abreviadamente la denotaremos como

$$V \hookrightarrow F_{n_1,n_2}(\delta) \ .$$

Igualmente, podría definirse una F de Snedecor doblemente descentrada donde las variables del numerador y del denominador son ambas χ^2 descentradas.

11.4 Ejercicios propuestos

11.1. Compruébese que si X sigue una distribución exponencial negativa de parámetro a, la variable $2aX$ es una χ^2 con 2 grados de libertad.

11.2. Demuéstrese que una distribución exponencial negativa de media 2 coincide con una distribución χ^2_2.

11.3. Hállese el intervalo intercuartílico de la distribución χ^2_{15}.

11.4. Obténgase la mediana de la distribución χ^2 con 10 grados de libertad.

11.5. Sea X una variable aleatoria con distribución χ^2. Estúdiese el crecimiento y decrecimiento de su función de densidad y determínese la moda.

11.6. Sean $X_1, \ldots, X_n$ variables aleatorias independientes e igualmente distribuidas con distribución $N(0, \sigma)$. Obténgase la distribución de la media de sus cuadrados:

$$Y = \frac{X_1^2 + \cdots + X_n^2}{n} \ .$$

11.7. Una niña coge todos los días el autobús para ir al colegio en una parada que está frente a su casa. El tiempo de espera diaria es una variable aleatoria exponencial negativa de media 5 minutos.
Su madre, convencida de que no es necesario que salga tan pronto de casa, decide hacer un análisis más detallado de la situación y anota los tiempos de espera de la niña durante 60 días. ¿Cuál es la probabilidad de que el tiempo medio de espera sea superior a 2 minutos?

11.8. Una modista tarda una media de 6 minutos en zurcir un calcetín. Sabiendo que el tiempo que emplea en realizar esta labor es una variable aleatoria con distribución gamma de parámetros a y p, siendo $p = 1$, calcúlese el porcentaje de calcetines en los que tarda más de 5 minutos.

11.9. Obténgase el quinto percentil y el séptimo decil de una distribución t de Student con 13 grados de libertad.

11.10. Sea X una variable con distribución t de Student con 20 grados de libertad. Hállese el intervalo real (a, b) más pequeño y el más grande para los cuales $p[X \in (a, b)] = 0.30$.

11.11. Sea X una variable aleatoria con distribución t de Student con n grados de libertad. Demuéstrese que $Y = X^2$ sigue una distribución F de Snedecor con 1 y n grados de libertad.

11.12. Sea X una variable aleatoria con distribución t de Student con n grados de libertad.

a) Calcúlese $E[|X|]$ para $n = 1$ y para $n > 1$.

b) Demuéstrese que, para $n > 1$, $E(X) = 0$.

11.13. Calcúlese el noveno decil de una distribución $F_{9,10}$.

11.14. Sea X una variable aleatoria con distribución F de Snedecor. Estúdiese el crecimiento y decrecimiento de su función de densidad y obténgase la moda.

11.15. Sean X_1 y X_2 variables aleatorias independientes con distribución $\gamma(a, p_1)$ y $\gamma(a, p_2)$, respectivamente, tales que $2p_1$ y $2p_2$ son números enteros. Hállese la distribución de $\frac{p_2 X_1}{p_1 X_2}$.

11.16. Sean $X_{11}, \ldots, X_{1n_1}$ y $X_{21}, \ldots, X_{2n_2}$ variables aleatorias independientes e igualmente distribuidas con distribución exponencial negativa. Sean las medias muestrales $\overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1i}}{n_1}$ y $\overline{X}_2 = \frac{\sum_{i=1}^{n_2} X_{2i}}{n_2}$. Obténgase la distribución de su cociente, $\frac{\overline{X}_1}{\overline{X}_2}$.

11.17. Sean $X_{11}, \ldots, X_{18}$ y $X_{21}, \ldots, X_{26}$, variables aleatorias independientes e igualmente distribuidas con distribución $N(0, 1)$.

a) Si $\overline{X}_1 = \frac{\sum_{i=1}^{8} X_{1i}}{8}$ y $\overline{X}_2 = \frac{\sum_{i=1}^{6} X_{2i}}{6}$, hállese el valor de la constante k tal que $p[\overline{X}_1 \geq k\overline{X}_2] = 0.10$.

b) Si $\overline{X}_{C_1} = \sqrt{\frac{\sum_{i=1}^{8} X_{1i}^2}{8}}$ y $\overline{X}_{C_2} = \sqrt{\frac{\sum_{i=1}^{6} X_{2i^2}}{6}}$, determínese el valor de k tal que $p[\overline{X}_{C_1} \geq k\overline{X}_{C_2}] = 0.10$.

11.18. ¿Qué condición debe cumplir una distribución F de Snedecor con n_1 y n_2 grados de libertad para que el m-ésimo percentil ($m = 1, \ldots, 99$) sea el inverso del $(100 - m)$-ésimo percentil?

11.19. Un médico cuya especialidad son los problemas de exceso de peso, somete a sus clientes a un nuevo tratamiento. La aplicación de dicho tratamiento en clínicas francesas determina que la proporción de pacientes en los que se produce una reducción de peso es una variable aleatoria beta de parámetros $p = 10$ y $q = 5$. Determínese la probabilidad de que tal proporción sea mayor que 0.8 para los clientes de este médico.

Capítulo 12

DISTRIBUCIÓN NORMAL n-DIMENSIONAL

Vamos a estudiar en este capítulo un importante modelo de distribución continua n-variante: la distribución normal n-dimensional. En un primer apartado desarrollaremos el caso general, concretando posteriormente los resultados para el caso bidimensional.

12.1 Distribución normal n-dimensional

Definiremos, en primer lugar, una variable normal n-dimensional para el caso en que sus componentes sean independientes. Más adelante, daremos la definición general de una variable normal n-dimensional, como resultado de realizar ciertas transformaciones lineales sobre n variables normales independientes.

Definición 12.1 *Sean $Y_1, \ldots, Y_n$, n variables aleatorias independientes e igualmente distribuidas con distribución $N(0,1)$. La variable aleatoria n-dimensional*

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

se dice que sigue una distribución normal n-dimensional estándar.

Puesto que las variables $Y_1, \ldots, Y_n$ son independientes, la función de densidad de la variable aleatoria n-dimensional $\mathbf{Y}$ es

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{Y_1}(y_1) \cdots f_{Y_n}(y_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}(y_1^2 + \cdots + y_n^2)} = \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\mathbf{y}'\mathbf{y}}, \end{aligned}$$

donde $-\infty < y_i < \infty$.

Definamos ahora una distribución normal n-dimensional cualquiera.

Definición 12.2 *Sea* $\mathbf{Y}$ *una variable aleatoria con distribución normal* n*-dimensional estándar,* $\mathbf{A} = (a_{ij})$ *una matriz cuadrada de orden* n *con determinante no nulo, y* $\mu = (\mu_1, \ldots, \mu_n)'$ *una matriz columna* $n \times 1$*. Sea* Σ *la matriz cuadrada de orden* n*,* $\Sigma = \mathbf{A} \cdot \mathbf{A}'$*. Entonces, si*

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{Y} + \mu \, ,$$

decimos que la variable aleatoria n*-dimensional*

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

sigue una distribución normal n*-dimensional de parámetros* μ *y* Σ*.*

Si $\mathbf{X}$ sigue una distribución normal n-dimensional de parámetros μ y Σ, lo expresaremos abreviadamente como

$$\mathbf{X} \hookrightarrow N_n(\mu, \Sigma) \, .$$

Evidentemente, si Y sigue una distribución normal n-dimensional estándar, $A = I$, matriz identidad de dimensión $n \times n$, y $\mu = \mathbf{0}$ matriz columna n-dimensional formada por ceros. Por tanto, los parámetros de una distribución normal n-dimensional estándar son $\mathbf{0}$ e I, denotándose abreviadamente por

$$\mathbf{Y} \hookrightarrow N_n(\mathbf{0}, I) \, .$$

En el estudio de variables normales unidimensionales, la proposición 10.4 indicaba que, a partir de una variable normal estándar, Z, es posible obtener una variable normal $N(\mu, \sigma)$, haciendo la transformación lineal, $X = \sigma \cdot Z + \mu$. La definición 12.2 muestra que las variables normales n-dimensionales se obtienen, también, mediante transformaciones lineales de una normal n-dimensional estándar. Si desarrollamos la anterior transformación,

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}Y_1 + \cdots + a_{1n}Y_n + \mu_1 \\ \vdots \\ a_{n1}Y_1 + \cdots + a_{nn}Y_n + \mu_n \end{pmatrix} .$$

En definitiva, para $i = 1, \ldots, n$,

$$X_i = a_{i1}Y_1 + \cdots + a_{in}Y_n + \mu_i \, ,$$

es decir, cada componente de una normal n-dimensional es una combinación lineal de n variables con distribución $N(0,1)$ e independientes. El hecho de que la matriz

A sea no singular, $|A| \neq 0$, implica que las componentes $X_1, \ldots, X_n$ son linealmente independientes, es decir, que ninguna de ellas se puede poner como combinación lineal del resto.

Calculemos la función de densidad de $\mathbf{X}$. Al ser la matriz $\mathbf{A}$ no singular, existe su inversa, con lo cual podemos escribir

$$\mathbf{Y} = \mathbf{A}^{-1} \cdot (\mathbf{X} - \mu) .$$

El jacobiano de este cambio de variable resulta ser[1]

$$J = |\mathbf{A}^{-1}| ,$$

y es no nulo; luego,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Y}}(\mathbf{y}(\mathbf{x})) \cdot |J| = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\left[(\mathbf{A}^{-1}(\mathbf{x}-\mu))'(\mathbf{A}^{-1}(\mathbf{x}-\mu))\right]} \cdot ||\mathbf{A}^{-1}|| = \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\left[(\mathbf{x}-\mu)'(\mathbf{A}^{-1})'\mathbf{A}^{-1}(\mathbf{x}-\mu)\right]} \cdot ||\mathbf{A}^{-1}|| . \end{aligned}$$

Simplifiquemos las expresiones que aparecen en la función de densidad anterior. Como $\Sigma = \mathbf{A} \cdot \mathbf{A}'$, se deduce que

$$\Sigma^{-1} = (\mathbf{A}^{-1})' \cdot \mathbf{A}^{-1} ,$$

y, por tanto,

$$|\Sigma^{-1}| = |(\mathbf{A}^{-1})'| \cdot |\mathbf{A}^{-1}| = (|\mathbf{A}^{-1}|)^2 .$$

Pero $|\Sigma^{-1}| = \frac{1}{|\Sigma|}$ y, en consecuencia, el valor absoluto de $|\mathbf{A}^{-1}|$ resulta

$$||A^{-1}|| = \frac{1}{\sqrt{|\Sigma|}} .$$

Lo anterior permite expresar la función de densidad de $\mathbf{X}$ en función de Σ:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\left[(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right]} ,$$

donde $-\infty < x_i < \infty$, para $i = 1, \ldots, n$.

La función de densidad de la normal n-dimensional $\mathbf{X} = \mathbf{A} \cdot \mathbf{Y} + \mu$ puede escribirse, por tanto, en función de dos matrices de parámetros, μ y $\Sigma = \mathbf{A}\mathbf{A}'$. La matriz $\mathbf{A}$ tiene sentido para describir la combinación lineal, pero las propiedades

1. El cambio de variable de $\mathbf{R}^n$ en $\mathbf{R}^n$ inducido por la transformación anterior es

$$\mathbf{y} = \mathbf{A}^{-1} \cdot (\mathbf{x} - \mu) .$$

Si llamamos b_{ij} $(i = 1, \ldots, n; j = 1, \ldots, n)$ a los elementos de $\mathbf{A}^{-1}$, y desarrollamos la expresión anterior, obtenemos

$$y_i = b_{i1}(x_1 - \mu_1) + \cdots + b_{in}(x_n - \mu_n).$$

Por tanto, $\frac{\partial y_i}{\partial x_j} = b_{ij}$ y, en definitiva, el jacobiano de la transformación es el determinante de la matriz $\mathbf{A}^{-1}$, esto es, $J = |\mathbf{A}^{-1}|$.

estadísticas de $\mathbf{X}$ dependen más bien de Σ que de $\mathbf{A}$. Como hemos visto, Σ se obtiene como el producto $\mathbf{A} \cdot \mathbf{A}'$. También es posible hallar $\mathbf{A}$ a partir de Σ, aunque el procedimiento no es sencillo, ya que implica diagonalizar esta última matriz (ejercicio propuesto 12.1).

Veamos, a continuación, las propiedades de la matriz Σ.

Proposición 12.1 *Σ es simétrica y definida positiva*[2].

Demostración. La matriz transpuesta de Σ coincide con Σ,

$$(\Sigma)' = (\mathbf{A}\mathbf{A}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{A}\mathbf{A}' = \Sigma\,,$$

y, por tanto, Σ es simétrica.

Comprobemos que, además, es definida positiva. Si $\mathbf{x}$ es un vector n-dimensional no nulo, veamos que $\mathbf{x}'\Sigma\mathbf{x}$ es siempre estrictamente positivo.

$$\mathbf{x}'\Sigma\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{A}'\mathbf{x} = (\mathbf{A}'\mathbf{x})'(\mathbf{A}'\mathbf{x})\,;$$

llamando $\mathbf{y} = \mathbf{A}'\mathbf{x}$,

$$\mathbf{x}'\Sigma\mathbf{x} = \mathbf{y}'\mathbf{y} = y_1^2 + \cdots + y_n^2 \geq 0\,.$$

Por tanto, $\mathbf{x}'\Sigma\mathbf{x} \geq 0$ para todo vector n-dimensional $\mathbf{x}$ y, en consecuencia, Σ es semidefinida positiva.

Además, la desigualdad anterior se convierte en igualdad solamente en el caso en que $y_1 = 0, \ldots, y_n = 0$, es decir, cuando $\mathbf{y} = \mathbf{0}$. Pero, entonces,

$$\mathbf{x} = (\mathbf{A}')^{-1}\mathbf{y} = (\mathbf{A}')^{-1}\mathbf{0} = \mathbf{0}\,,$$

lo que implica que Σ es definida positiva. ♣

Estas propiedades se trasladan de una matriz a su inversa. Veámoslo para la matriz Σ.

Proposición 12.2 *Σ^{-1} es simétrica y definida positiva.*

Demostración. Probaremos que la inversa de cualquier matriz simétrica y definida positiva también cumple ambas propiedades.

Sea $\mathbf{B}$ una matriz simétrica y definida positiva. Como B es simétrica, $\mathbf{B}' = \mathbf{B}$; por tanto,

$$(\mathbf{B}^{-1})' = (\mathbf{B}')^{-1} = \mathbf{B}^{-1}\,,$$

con lo que $\mathbf{B}^{-1}$ es simétrica.

Veamos que $\mathbf{B}^{-1}$ es definida positiva. Recuerde el lector que una matriz es definida positiva si, y sólo si, sus valores propios son todos estrictamente positivos. En consecuencia, todos los valores propios de $\mathbf{B}$ serán positivos. Pero los valores propios de $\mathbf{B}^{-1}$ son los inversos de los de $\mathbf{B}$ y, consecuentemente, serán, también, positivos. En definitiva, $\mathbf{B}^{-1}$ es una matriz definida positiva.

De lo anterior deducimos que Σ^{-1} es simétrica y definida positiva por serlo Σ. ♣

2. Advierta el lector que este enunciado es redundante, ya que toda matriz definida positiva ha de ser, por definición, simétrica. El motivo de enunciar así la proposición no es otro que el de recalcar los conceptos.

Los resultados anteriores van a permitir definir una normal n-dimensional de manera más operativa. Sabemos que su función de densidad puede expresarse en función de μ y de Σ, donde μ puede ser cualquiera, pero Σ es una matriz que tiene que poderse escribir como producto de una matriz cuadrada no singular, $\mathbf{A}$, por su transpuesta. Hemos visto en las dos proposiciones anteriores que si una matriz se puede escribir así, es simétrica y definida positiva. El lector puede probar que, recíprocamente, si una matriz es simétrica y definida positiva, admite una descomposición de ese tipo (ejercicio propuesto 12.1). En definitiva, ambas propiedades caracterizan a las matrices Σ. La siguiente definición formaliza este comentario, permitiendo definir una normal n-dimensional sin necesidad de hacer referencia a la normal n-dimensional estándar.

Definición 12.3 *La variable aleatoria n-dimensional $\mathbf{X} = (X_1, \ldots, X_n)'$ sigue una distribución normal n-dimensional de parámetros μ y Σ, con $\mu = (\mu_1, \ldots, \mu_n)'$ $(-\infty < \mu_i < \infty$, para $i = 1, \ldots, n)$ y $\Sigma = (\sigma_{ij})$, matriz simétrica y definida positiva, si su función de densidad es*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}[(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)]},$$

para $-\infty < x_i < \infty$ $(i = 1, \ldots, n)$.

El lector puede comprobar que, para $n = 1$, la función de densidad coincide con la de la normal unidimensional.

Aunque, como dijimos en la sección 10.2, los estudios acerca de las distribuciones normales unidimensionales se desarrollan en los comienzos del siglo XVIII, hasta los primeros años del siglo XIX no se estudian distribuciones conjuntas de dos variables normales, es decir, lo que hoy día se entiende como una distribución normal bivariante. Ejemplos de ello son los trabajos de R. ADRIAN en 1808 o A. BRAVAIS en 1846 sobre probabilidades de errores de medida. Es, sin embargo, a finales de siglo y, concretamente, en su último cuarto, cuando se formaliza la definición de la distribución normal bivariante. F. GALTON aplica esta distribución al estudio de la herencia genética, y sugiere a I. D. H. DICKSON la posibilidad de definir la distribución de una variable normal bidimensional, combinando variables normales (unidimensionales) independientes.

En el primer lustro de nuestro siglo, K. PEARSON realiza un estudio bastante completo de la distribución normal bidimensional dentro de sus trabajos biométricos sobre la teoría de la evolución, e incluso inicia la tabulación de su función de distribución. Las extensión a más de dos dimensiones es posterior. La propiedad de las variables normales que indica que la incorrelación entre las variables implica su independencia, fue el motivo que en su día dificultó los primeros intentos de generalización a dimensiones superiores a dos, al tratar de generalizar un concepto, el de incorrelación, que sólo tiene sentido para pares de variables.

La función de distribución de una normal n-dimensional no tiene una expresión explícita, al no conocerse primitiva de la función de densidad. Para calcular probabilidades, existen diversas aproximaciones que, para $n = 2, 3, 4$, se concretan en tablas que recogen los valores de la función de distribución.

Las matrices μ y Σ que aparecen en la función de densidad de una $N_n(\mu, \Sigma)$ tienen un significado estadístico importante, ya que los elementos de μ son las esperanzas de las variables $X_1, \ldots, X_n$, y los de Σ, las varianzas de dichas variables y las covarianzas entre cada par de ellas. Es decir, μ es la esperanza de $\mathbf{X}$ y Σ es su matriz de varianzas-covarianzas. Veámoslo, en primer lugar, para una normal n-dimensional estándar y, más adelante, para una normal n-dimensional cualquiera.

Proposición 12.3 *Sea* $\mathbf{Y}$ *una variable aleatoria n-dimensional con distribución* $N_n(\mathbf{0}, \mathbf{I})$. *Entonces, la esperanza de* $\mathbf{Y}$ *es la matriz columna* $\mathbf{0}$, *aquella en la que todos sus elementos valen cero, y su matriz de varianzas-covarianzas es la matriz identidad* $\mathbf{I}$.

Demostración. Las variables $Y_1, \ldots, Y_n$ son $N(0,1)$; por tanto, $E(Y_i) = 0$ para $i = 1, \ldots, n$ y, consecuentemente,

$$E(\mathbf{Y}) = \mathbf{0}\,.$$

Además, $Var(Y_i) = 1$ para $i = 1, \ldots, n$. Por otra parte, las variables $Y_1, \ldots, Y_n$ son independientes; estarán, pues, incorrelacionadas y $Cov(Y_i, Y_j) = 0$ para $i \neq j$. Luego la matriz de varianzas-covarianzas tiene unos en la diagonal principal y ceros fuera de la misma, es decir,

$$V(\mathbf{Y}) = \mathbf{I}\,.\clubsuit$$

Proposición 12.4 *Sea* $\mathbf{X}$ *una variable aleatoria n-dimensional con distribución* $N_n(\mu, \Sigma)$. *Entonces, la esperanza de* $\mathbf{X}$ *es la matriz columna* μ *y su matriz de varianzas-covarianzas es* Σ.

Demostración. Puesto que, como ya señalamos, $\mathbf{X} = \mathbf{A} \cdot \mathbf{Y} + \mu$, su esperanza resulta

$$E(\mathbf{X}) = \mathbf{A} \cdot E(\mathbf{Y}) + \mu = \mathbf{A} \cdot \mathbf{0} + \mu = \mu\,.$$

Además, de las propiedades de la matriz de varianzas-covarianzas se deduce que

$$V(\mathbf{X}) = V(\mathbf{AY} + \mu) = \mathbf{A}V(\mathbf{Y})\mathbf{A}' = \mathbf{AIA}' = \mathbf{AA}' = \Sigma\,.\clubsuit$$

Obtengamos, a continuación, la función característica de una normal n-dimensional, que va a permitir demostrar, de forma sencilla, importantes propiedades.

Proposición 12.5 *Sea* $\mathbf{X}$ *una variable aleatoria n-dimensional con distribución* $N_n(\mu, \Sigma)$. *Su función característica es*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}\,,$$

donde $\mathbf{t} = (t_1, \ldots, t_n)'$.

Demostración. Sea $\mathbf{Y} \hookrightarrow N_n(\mathbf{0}, \mathbf{I})$; entonces, $Y_1, \ldots, Y_n$ son independientes. Por tanto,

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \varphi_{Y_1}(t_1) \cdots \varphi_{Y_n}(t_n) = e^{-\frac{t_1^2}{2}} \cdots e^{-\frac{t_n^2}{2}} = e^{-\frac{1}{2}(t_1^2 + \cdots + t_n^2)} = e^{-\frac{1}{2}\mathbf{t}'\mathbf{t}} .$$

Teniendo en cuenta que si $\mathbf{X}$ sigue una distribución normal n-dimensional $N_n(\mu, \Sigma)$, se puede expresar en función de $\mathbf{Y}$ como $\mathbf{X} = \mathbf{AY} + \mu$,

$$\begin{aligned} \varphi_{\mathbf{X}}(\mathbf{t}) &= E\left(e^{i\mathbf{t}'\mathbf{X}}\right) = E\left(e^{i\mathbf{t}'\mathbf{AY}+\mu}\right) = e^{i\mathbf{t}'\mu} E(e^{i\mathbf{t}'\mathbf{AY}}) = \\ &= e^{i\mathbf{t}'\mu} \cdot \varphi_{\mathbf{Y}}(\mathbf{A}'\mathbf{t}) = e^{i\mathbf{t}'\mu} \cdot e^{-\frac{1}{2}\mathbf{t}'\mathbf{A}\mathbf{A}'\mathbf{t}} = \\ &= e^{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}} . \clubsuit \end{aligned}$$

Veamos que, en muchos casos, si hacemos una transformación lineal de una variable normal n-dimensional, el resultado es, igualmente, una variable del mismo tipo.

Proposición 12.6 *Sea* $\mathbf{X}$ *una variable aleatoria* n*-dimensional con distribución* $N_n(\mu, \Sigma)$*, y sean dos matrices de números reales,* $\mathbf{B}$*, de dimensión* $p \times n$*, y* $\mathbf{C}$*, de dimensión* $p \times 1$*. Entonces, la variable aleatoria* p*-dimensional* $\mathbf{Z} = \mathbf{B} \cdot \mathbf{X} + \mathbf{C}$ *tiene una distribución* $N_p(\mathbf{B} \cdot \mu + \mathbf{C}, \mathbf{B} \cdot \Sigma \cdot \mathbf{B}')$*, siempre que* $\mathbf{B}\Sigma\mathbf{B}'$ *sea una matriz definida positiva.*

Demostración. Evidentemente, $\mathbf{Z}$ es una variable aleatoria p-dimensional; calculemos su función característica,

$$\begin{aligned} \varphi_{\mathbf{Z}}(\mathbf{t}) &= E\left(e^{i\mathbf{t}'\mathbf{Z}}\right) = E\left(e^{i\mathbf{t}'(\mathbf{BX}+\mathbf{C})}\right) = e^{i\mathbf{t}'\mathbf{C}} \cdot E\left(e^{i\mathbf{t}'\mathbf{BX}}\right) = \\ &= e^{i\mathbf{t}'\mathbf{C}} \cdot \varphi_{\mathbf{X}}(\mathbf{B}'\mathbf{t}) = e^{i\mathbf{t}'\mathbf{C}} \cdot e^{i\mathbf{t}'\mathbf{B}\mu - \frac{1}{2}\mathbf{t}'\mathbf{B}\Sigma\mathbf{B}'\mathbf{t}} = \\ &= e^{i\mathbf{t}'(\mathbf{B}\mu+\mathbf{C}) - \frac{1}{2}\mathbf{t}'\mathbf{B}\Sigma\mathbf{B}'\mathbf{t}} , \end{aligned}$$

que es la función característica de una normal p-dimensional, $N_p(\mathbf{B} \cdot \mu + \mathbf{C}, \mathbf{B} \cdot \Sigma \cdot \mathbf{B}')$, siempre que $\mathbf{B} \cdot \Sigma \cdot \mathbf{B}'$ pueda ser su matriz de varianzas-covarianzas, esto es, siempre que sea simétrica y definida positiva. $\clubsuit$

Obsérvese que la variable $\mathbf{Z} = \mathbf{B} \cdot \mathbf{X} + \mathbf{C}$ recoge p transformaciones lineales de las variables $X_1, \ldots, X_n$, a las que se ha añadido un término constante. Desarrollemos la expresión matricial que define la variable $\mathbf{Z}$,

$$\begin{aligned} \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix} &= \begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & & \vdots \\ b_{p1} & \ldots & b_{pn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix} = \\ &= \begin{pmatrix} b_{11}X_1 & + \cdots + & b_{1n}X_n + c_1 \\ \vdots & & \vdots \\ b_{p1}X_1 & + \cdots + & b_{pn}X_n + c_p \end{pmatrix} . \end{aligned}$$

En consecuencia, las variables

$$Z_i = b_{i1}X_1 + \cdots + b_{in}X_n + c_i \ , \quad i = 1, \ldots, p \ ,$$

son combinaciones lineales de las componentes de la variable normal n-dimensional.

Veamos qué condiciones debe cumplir $\mathbf{B}$ para que $\mathbf{B\Sigma B'}$ sea simétrica y definida positiva, esto es, para que la variable transformada, $\mathbf{Z}$, sea normal p-dimensional, de acuerdo con la proposición anterior.

La matriz $\mathbf{B\Sigma B'}$ es la matriz de varianzas-covarianzas de la nueva variable $\mathbf{Z}$, pues

$$V(\mathbf{Z}) = V(\mathbf{BX} + \mathbf{C}) = \mathbf{B}V(\mathbf{X})\mathbf{B'} = \mathbf{B\Sigma B'} \ .$$

Por ser una matriz de varianzas-covarianzas ha de ser siempre simétrica y semidefinida positiva, como vimos en el apartado 7.5, sea cual sea $\mathbf{B}$.

En este mismo apartado vimos que la matriz de varianzas-covarianzas es definida positiva si, y sólo si, no existe ninguna combinación lineal entre las variables. Entonces, tenemos que exigir que entre las componentes de $\mathbf{Z} = \mathbf{BX} + \mathbf{C}$, $Z_1, \ldots, Z_p$, no exista ninguna combinación lineal. Esto es equivalente a exigir que $\mathbf{B}$ sea una matriz no singular, o lo que es lo mismo, que el rango de $\mathbf{B}$ sea p (obviamente, entre las componentes de $\mathbf{X}$ tampoco existe ninguna combinación lineal, pues Σ es, también, definida positiva).

Pero, como $\mathbf{B}$ es una matriz de dimensión $p \times n$, la condición impuesta a su rango tiene dos consecuencias importantes:

1. El número de filas, p, no puede rebasar el de columnas, n ($p \not> n$). Esto es, si hacemos una cantidad, p, mayor que n, de combinaciones lineales de las componentes de una normal n-dimensional, la variable resultante no será una normal p-dimensional.
2. Si $p \leq n$, las filas de $\mathbf{B}$ tienen que ser linealmente independientes para que su rango sea p.

Es decir, si tenemos una variable aleatoria normal de dimensión, por ejemplo, 4, (X_1, X_2, X_3, X_4), por transformaciones lineales de la misma podemos encontrar normales de dimensión 4, 3, 2 o 1, pero nunca de dimensión 5 o mayor. En el caso en que, a partir de (X_1, X_2, X_3, X_4) construyamos una variable, por ejemplo, tridimensional, (Z_1, Z_2, Z_3), para que ésta sea una normal tenemos que exigir que ninguna de las variables Z_i $(i = 1, 2, 3)$ se pueda poner como combinación lineal del resto.

Como consecuencias directas de la proposición 12.6, estudiemos los tres corolarios siguientes.

Corolario 12.6.1 *Sea* $\mathbf{X}$ *una variable aleatoria* n*-dimensional con distribución* $N_n(\mu, \Sigma)$*, y sea* $\mathbf{B}$ *una matriz de números reales* $p \times n$ *de rango* p*; entonces,* $\mathbf{Z} = \mathbf{B} \cdot \mathbf{X}$ *sigue una distribución* $N_p(\mathbf{B} \cdot \mu, \mathbf{B} \cdot \Sigma \cdot \mathbf{B'})$.

La demostración es inmediata haciendo $\mathbf{C} = \mathbf{0}$ en la proposición anterior.

Ejemplo 12.1. En un cierto país, la renta y el consumo mensual (en miles de pesetas) de una familia, (X_1, X_2), sigue una distribución normal bidimensional de medias $\mu_1 = 100$ y $\mu_2 = 90$, y de varianzas $\sigma_{11} = 900$ y $\sigma_{22} = 625$. Sabiendo que sólo el 10 % de las familias se endeuda mensualmente, ¿qué porcentaje de familias ahorra al menos el 20 % de su renta mensual?

Para que una familia ahorre al menos el 20 % de sus ingresos, es necesario que la diferencia entre la renta, X_1, y el consumo, X_2, sea mayor o igual que el 20 % de X_1; es decir, hemos de calcular:

$$p[X_1 - X_2 \geq 0.2X_1] = p[0.8X_1 - X_2 \geq 0] ,$$

para lo cual necesitamos conocer la distribución de la variable $V = 0.8X_1 - X_2$.

Sabemos que

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \hookrightarrow N_2(\mu, \Sigma) ,$$

donde $\mu = \binom{100}{90}$ y $\Sigma = \begin{pmatrix} 900 & \sigma_{12} \\ \sigma_{21} & 625 \end{pmatrix}$, con $\sigma_{12} = \sigma_{21} = Cov(X_1, X_2)$.

Expresemos V en forma matricial. Así,

$$V = 0.8X_1 - X_2 = (0.8, -1)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mathbf{BX} ,$$

donde $\mathbf{B} = (0.8, -1)$. Aplicando el corolario 12.6.1, deducimos que V es normal unidimensional de media

$$\mathbf{B}\mu = (0.8, -1)\begin{pmatrix} 100 \\ 90 \end{pmatrix} = 0.8 \cdot 100 - 90 = -10$$

y varianza

$$\mathbf{B}\Sigma\mathbf{B}' = (0.8, -1)\begin{pmatrix} 900 & \sigma_{12} \\ \sigma_{12} & 625 \end{pmatrix}\begin{pmatrix} 0.8 \\ -1 \end{pmatrix} = 1201 - 1.6\sigma_{12} .$$

Por tanto,

$$V \hookrightarrow N\left(-10, \sqrt{1201 - 1.6\sigma_{12}}\right) .$$

Antes de continuar con la obtención de la probabilidad referida a la variable V, conviene mencionar que el corolario 12.6.1 permite asegurar que V tiene una distribución normal; por ello, y al tratarse de una variable unidimensional, bastaría conocer su media y varianza para determinar completamente su distribución (obsérvese que esto no sería válido si V fuese p-dimensional). Así,

$$E(V) = E(0.8X_1 - X_2) = 0.8E(X_1) - E(X_2) = -10$$

y

$$\begin{aligned} Var(V) &= Var(0.8X_1 - X_2) = \\ &= 0.8^2 Var(X_1) + Var(X_2) - 2 \cdot 0.8 Cov(X_1, X_2) = 1201 - 1.6\sigma_{12} . \end{aligned}$$

Siguiendo con el cálculo de la probabilidad que nos ocupa, habrá que hallar σ_{12} y determinar, de este modo, el valor de la varianza de V. Para ello, utilizaremos el dato de que el 10 % de las familias se endeuda, es decir, el 10 % tiene un consumo superior a su renta, esto es,

$$p[X_1 < X_2] = p[X_1 - X_2 < 0] = 0.10 .$$

Definiendo $U = X_1 - X_2$, y teniendo en cuenta el corolario anterior, con $B = (1, -1)$, la variable

$$U = (1, -1)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

sigue una distribución

$$N_1\left[(1,-1)\begin{pmatrix}100\\90\end{pmatrix}, (1,-1)\begin{pmatrix} 900 & \sigma_{12} \\ \sigma_{12} & 625 \end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix}\right] \equiv N_1(10, \sqrt{1525 - 2\sigma_{12}}) .$$

Sabemos que

$$p[X_1 < X_2] = p[U < 0] = \Phi\left(\frac{0-10}{\sqrt{1525 - 2\sigma_{12}}}\right) = 0.1 .$$

Buscando en las tablas de la distribución $N(0,1)$, tendremos que

$$\frac{-10}{\sqrt{1525 - 2\sigma_{12}}} = -1.28 ,$$

que, despejando, proporciona el valor de $\sigma_{12} = 731.98$. Según esto,

$$V \hookrightarrow N(-10, \sqrt{1201 - 1.6 \cdot 731.98}) \equiv N(-10, 5.4619) .$$

Por tanto, estamos en condiciones de calcular la probabilidad pedida:

$$\begin{aligned} p[X_1 - X_2 \geq 0.2X_1] &= p[0.8X_1 - X_2 \geq 0] = p[V \geq 0] = 1 - p[V < 0] = \\ &= 1 - \Phi\left(\frac{10}{5.4619}\right) = 1 - \Phi(1.83) = 0.03362 . \end{aligned}$$

Es decir, sólo el 3.36 % de las familias ahorra más del 20 % de sus ingresos. ◁

Corolario 12.6.2 *Sea* $\mathbf{X}$ *una variable aleatoria* n*-dimensional con distribución* $N_n(\mu, \Sigma)$. *La variable aleatoria formada por cualquier subconjunto de* k *variables de las* n *iniciales, sigue una distribución normal* k*-dimensional.*

Demostración. Para simplificar, sin que ello provoque una pérdida de generalidad, tomemos las k primeras variables de $\mathbf{X}$, y denotemos por $\mathbf{X}^*$ a la variable aleatoria formada por dichas variables, $\mathbf{X}^* = (X_1, \ldots, X_k)'$. Llamando $\mathbf{B}$ a la matriz $k \times n$ $(k < n)$

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{pmatrix} ,$$

de la proposición 12.6 se deduce que

$$\mathbf{X}^* = \mathbf{B} \cdot \mathbf{X} \hookrightarrow N_k(\mu^*, \Sigma^*) ,$$

donde $\mu^* = \mathbf{B} \cdot \mu = (\mu_1, \ldots, \mu_k)'$ y $\Sigma^* = \mathbf{B}\Sigma\mathbf{B}'$ es la submatriz $k \times k$ de Σ formada por los elementos de sus k primeras filas y columnas. ♣

Obsérvese que, como en el ejemplo anterior, la afirmación nueva que encierra este corolario es la normalidad de $\mathbf{X}^*$, ya que los elementos de μ^* y Σ^* eran conocidos de antemano, al tratarse de las esperanzas de las componentes de $\mathbf{X}^*$ y de sus varianzas y covarianzas.

Corolario 12.6.3 *Sea* $\mathbf{X}$ *una variable aleatoria* n*-dimensional con distribución* $N_n(\mu, \Sigma)$. *Entonces, la variable aleatoria unidimensional* X_i $(i = 1, \ldots, n)$ *tiene una distribución* $N_1(\mu_i, \sqrt{\sigma_{ii}})$.

Demostración. Es evidente, sin más que considerar en el corolario anterior la matriz $\mathbf{B} = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ de dimensión $1 \times n$, cuyos valores son todos nulos salvo el elemento i-ésimo, que es igual a la unidad. ♣

El corolario anterior demuestra, por tanto, que las marginales de una normal n-dimensional son normales. El siguiente ejemplo ilustra dicho resultado.

Ejemplo 12.2. En la situación del ejemplo 12.1, ¿qué porcentaje de familias tiene un consumo mensual superior a 150 mil pesetas?

Por el corolario 12.6.3, las marginales de una distribución normal n-dimensional son normales unidimensionales; en particular, el consumo mensual, X_2, sigue una distribución $N(90, \sqrt{625})$. Así,

$$p[X_2 > 150] = 1 - p[X_2 \leq 150] = 1 - \Phi\left(\frac{150 - 90}{25}\right) = 1 - \Phi(2.4) = 0.008198\,.$$

Es decir, solamente un 0.82 % de las familias tiene una renta mensual superior a 150 mil pesetas. ◁

El resultado recíproco del corolario anterior no es cierto. Esto es, si las distribuciones marginales de una variable n-dimensional son normales, su distribución conjunta no tiene por qué ser una normal n-dimensional. En el fondo, lo que ocurre es que distintas distribuciones conjuntas pueden proporcionar las mismas distribuciones marginales. Veámoslo con el siguiente ejemplo.

Ejemplo 12.3. Sea la variable bidimensional continua (X, Y), con función de densidad

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{\pi} \cdot e^{-\frac{1}{2}(x^2+y^2)} & \begin{cases} x < 0, y > 0 \\ x > 0, y < 0 \end{cases} \\ 0 & \text{en el resto} \end{cases}$$

Obviamente, esta distribución conjunta no es normal bidimensional (esta variable toma sus valores exclusivamente en los cuadrantes segundo y cuarto del plano, mientras que una normal bidimensional varía en todo el plano real). Sin embargo, puede comprobar el lector que las distribuciones marginales son normales unidimensionales; de hecho, son $N(0, 1)$. Por tanto, las distribuciones marginales son normales no siéndolo la conjunta.

Ahora bien, si calculamos su covarianza, ésta vale $-\frac{2}{\pi}$. Su matriz de varianzas-covarianzas es, por tanto,

$$\Sigma = \begin{pmatrix} 1 & -\frac{2}{\pi} \\ -\frac{2}{\pi} & 1 \end{pmatrix}.$$

La distribución conjunta anterior no es la única que tiene las dos marginales $N(0, 1)$. También se obtendrían las mismas distribuciones marginales si la distribución conjunta fuese una normal bidimensional, en concreto,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \hookrightarrow N_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -\frac{2}{\pi} \\ -\frac{2}{\pi} & 1 \end{pmatrix}\right]. \quad \triangleleft$$

Si las distribuciones marginales son independientes, sabemos que la función de densidad conjunta queda determinada de forma única por las marginales, ya que es el producto de éstas. Si, además, son normales, queda garantizada la normalidad de la distribución conjunta.

Proposición 12.7 *Sean $X_1, \ldots, X_n$, n variables aleatorias independientes con distribución $N(\mu_i, \sqrt{\sigma_{ii}})$. La variable aleatoria $\mathbf{X}$ formada por las n variables sigue una distribución normal n-dimensional.*

Demostración. Si las variable $X_1, \ldots, X_n$ son independientes, su función de densidad conjunta será el producto de las marginales, esto es,

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{ii}}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{ii}}} =$$

$$= \frac{1}{\sqrt{\sigma_{11}\cdots\sigma_{nn}}(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}[(\mathbf{X}-\mu)'\cdot\Sigma^{-1}\cdot(\mathbf{X}-\mu)]},$$

que es la función de densidad de una $N_n(\mu, \Sigma)$, con $\mu = (\mu_1, \ldots, \mu_n)'$ y

$$\Sigma = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{nn} \end{pmatrix} . \clubsuit$$

Existe una caracterización de la normalidad de una variable n-dimensional a través del estudio de variables aleatorias de dimensión igual a la unidad, que, a veces, se podrá utilizar con éxito. La clave de esta caracterización no son las distribuciones marginales, sino las de combinaciones lineales de las variables.

Proposición 12.8 *Sea $\mathbf{X}$ una variable aleatoria n-dimensional, con esperanza $E(\mathbf{X}) = \mu$, y matriz de varianzas-covarianzas $V(\mathbf{X}) = \Sigma$, simétrica y definida positiva. La variable aleatoria $\mathbf{X}$ sigue una distribución normal n-dimensional si, y sólo si, cualquier combinación lineal de sus componentes sigue una distribución normal unidimensional; es decir,*

$$\mathbf{X} \hookrightarrow N_n(\mu, \Sigma) \iff \sum_{i=1}^{n} b_i X_i \hookrightarrow N_1\left(\sum_{1}^{n} b_i\mu_i, \sqrt{\sum_{i,j=1}^{n} b_i b_j \sigma_{ij}}\right),$$

para cualesquiera números reales $b_1, \ldots, b_n$ no todos nulos.

Demostración. Para probar la condición necesaria, consideremos la matriz de números reales $\mathbf{B} = (b_1, \ldots, b_n)$, donde algún b_i es no nulo. Como $\sum_{i=1}^{n} b_i X_i = \mathbf{B} \cdot \mathbf{X}$, en virtud del corolario 12.6.1 se tiene que su distribución es

$$N_1(\mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}') \equiv N_1\left(\sum_{1}^{n} b_i\mu_i, \sqrt{\sum_{i,j=1}^{n} b_i b_j \sigma_{ij}}\right),$$

donde la varianza de $\mathbf{BX}$, $\mathbf{B}\Sigma\mathbf{B}'$, es $\sum_{i,j=1}^{n} b_i b_j \sigma_{ij}$.

Para demostrar la condición suficiente, podemos expresar la función característica de $\sum_{i=1}^{n} b_i X_i = \mathbf{B} \cdot \mathbf{X}$ como

$$\varphi_{\mathbf{BX}}(\mathbf{t}) = E(e^{i\mathbf{tBX}}) = \varphi_{\mathbf{X}}(\mathbf{t} \cdot \mathbf{B}') .$$

Puesto que $\mathbf{BX}$, por hipótesis, es una normal unidimensional de media $\mathbf{B}\mu$ y varianza $\mathbf{B}\Sigma\mathbf{B}'$, resulta que su función característica es

$$\varphi_{\mathbf{BX}}(t) = e^{it\mathbf{B}\mu - \frac{1}{2}t^2\mathbf{B}\Sigma\mathbf{B}'} ,$$

que corresponde, también, a la función característica de $\mathbf{X}$ en el punto $t\mathbf{B}'$. Haciendo $t = 1$ resulta:

$$\varphi_{\mathbf{X}}(\mathbf{B}') = e^{i\mathbf{B}\mu - \frac{1}{2}\mathbf{B}\Sigma\mathbf{B}'} ,$$

siendo $\mathbf{B} = (b_1, \ldots, b_n)$ con no todas las componentes nulas. Esta función característica es la de una normal $N_n(\mu, \Sigma)$, con lo cual queda demostrada la condición suficiente. ♣

Recordemos que en la proposición 10.5 ya demostramos que cualquier combinación lineal de variables normales independientes sigue una distribución normal. Cuando las variables normales provienen de una normal n-dimensional no hay, pues, que exigir la independencia.

Como se sabe, si varias variables son independientes están incorrelacionadas, esto es, el coeficiente de correlación entre dos cualesquiera de ellas es igual a cero. No obstante, como vimos en el apartado 7.1, en general, no se cumple el resultado recíproco, es decir, varias variables pueden estar incorrelacionadas y, sin embargo, ser dependientes. La siguiente proposición demuestra que, para variables normales n-dimensionales, ambos conceptos son equivalentes.

Proposición 12.9 *Sea* $\mathbf{X}$ *una variable aleatoria* n*-dimensional con distribución* $N_n(\mu, \Sigma)$. *Las* n *componentes de la variable aleatoria* n*-dimensional* $\mathbf{X}$, $X_1, \ldots, X_n$, *son independientes si, y sólo si, están incorrelacionadas.*

Demostración. Como vimos en el apartado 7.1, si varias variables son independientes, están incorrelacionadas. Comprobemos que, para variables normales n-dimensionales, el recíproco también es cierto.

Si $X_1, \ldots, X_n$ son variables aleatorias incorrelacionadas, la inversa de su matriz de varianzas-covarianzas es

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{nn}} \end{pmatrix} ,$$

por lo que el producto $(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)$ resulta ser igual a

$$(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_{ii}} .$$

Deducimos, así, que la función de densidad de $\mathbf{X}$ es

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{|\Sigma|}(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}[(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)]} = \\ &= \frac{1}{\sqrt{\prod \sigma_{ii}}(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\left[\sum_{i=1}^{n} \frac{(X_i-\mu_i)^2}{\sigma_{ii}}\right]} = \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma_{ii}2\pi}} \cdot e^{-\frac{(X_i-\mu_i)^2}{2\sigma_{ii}}} = f_{X_1}(x_1)\cdots f_{X_n}(x_n) . \end{aligned}$$

Como hemos demostrado que la función de densidad conjunta es el producto de las funciones de densidad marginales, $X_1, \ldots, X_n$ son independientes. ♣

Veamos, finalmente, una proposición que indica que, si se conocen los valores que toman varias de las componentes de una normal n-dimensional, la distribución condicionada de las restantes componentes es también normal. Omitiremos la demostración del resultado.

Sea $\mathbf{X}$ una variable aleatoria con distribución $N_n(\mu, \Sigma)$. Dividamos sus componentes en grupos; con el superíndice (1) denotaremos a las p primeras componentes y con el superíndice (2) las restantes. Así,

$$\begin{aligned} \mathbf{X} &= (X_1, \ldots, X_p, X_{p+1}, \ldots, X_n)' = (X^{(1)}, X^{(2)})' , \\ \mu &= (\mu_1, \ldots, \mu_p, \mu_{p+1}, \ldots, \mu_n)' = (\mu^{(1)}, \mu^{(2)})' \quad \text{y} \\ \Sigma &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} , \end{aligned}$$

donde las submatrices Σ_{ij} se obtienen dividiendo la matriz Σ en submatrices de dimensión $p \times p$, $p \times n-p$, $n-p \times p$ y $n-p \times n-p$, respectivamente.

Enunciemos en estos términos el resultado descrito en el comentario anterior.

Proposición 12.10 *La distribución de $X^{(1)}$ condicionada por $X^{(2)}$ es una normal p-dimensional de media $\mu^{(1)} + \Sigma_{12} \cdot \Sigma_{22}^{-1}(X^{(2)} - \mu^{(2)})$ y matriz de varianzas-covarianzas $\Sigma_{11} - \Sigma_{12} \cdot \Sigma_{22}^{-1} \cdot \Sigma_{21}$.*

12.2 Distribución normal bidimensional

Esta distribución es un caso particular de la distribución normal n-dimensional para $n = 2$. En consecuencia, son válidos todos los resultados y propiedades obtenidos en este capítulo.

Dedicamos a su estudio un apartado por varios motivos. En primer lugar, por motivos históricos ya que, como ya indicamos, fue la primera distribución continua multivariante estudiada. Pero, además, hoy en día es muy frecuente que, cuando se modelizan fenómenos aleatorios bivariantes, asignemos a las variables esta distribución.

Finalmente, su desarrollo resulta ilustrativo de los resultados demostrados para las variables normales n-dimensionales, ya que pueden escribirse sin recurrir a la

notación matricial, evidenciando, de forma explícita, el contenido de dichos resultados. Comencemos definiendo una normal bidimensional estándar.

Definición 12.4 *Sean X e Y variables aleatorias independientes e igualmente distribuidas con distribución $N(0,1)$. Decimos, entonces, que la variable aleatoria bidimensional (X,Y) sigue una distribución normal bidimensional estándar.*

Si (X,Y) sigue una distribución normal estándar, lo expresaremos abreviadamente como

$$(X,Y) \hookrightarrow N_2(\mathbf{0},\mathbf{I})$$

donde $\mathbf{0} = (0,0)'$ e $\mathbf{I}$ es la matriz identidad de orden dos.

Como X e Y son independientes, la función de densidad de (X,Y) resulta:

$$f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} ,$$

para $-\infty < x < \infty$ y $-\infty < y < \infty$.

De igual modo, puede definirse una distribución normal bidimensional general. Si concretásemos la definición 12.2, sería el resultado de una transformación lineal a partir de una normal bidimensional estándar. Con todo, en este caso es más habitual definirla directamente.

Definición 12.5 *Sea (X,Y) una variable aleatoria bidimensional. Decimos que (X,Y) sigue una distribución normal bidimensional de parámetro μ y Σ, con $\mu = (\mu_1,\mu_2)'$ ($-\infty < \mu_i < \infty$ para $i = 1,2$) y Σ matriz 2×2 simétrica y definida positiva, si su función de densidad es*

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi \cdot \sqrt{|\Sigma|}} \cdot \exp\left\{ -\frac{1}{2} \left[(x-\mu_1, y-\mu_2) \Sigma^{-1} \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix} \right] \right\} ,$$

para $-\infty < x < \infty$ y $-\infty < y < \infty$.

Como demostramos en la proposición 12.4, μ es la matriz de medias de (X,Y) y Σ es su matriz de varianzas-covarianzas. Por tanto, las componentes de μ serán las esperanzas de las variables X e Y, μ_X y μ_Y, esto es, $\mu = (\mu_X, \mu_Y)'$.

Las componentes de Σ son las varianzas y la covarianza de ambas variables, es decir, $\sigma_{11} = \sigma_X^2$, $\sigma_{22} = \sigma_Y^2$, y $\sigma_{12} = \sigma_{21} = \sigma_{XY}$. En definitiva,

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} .$$

No obstante, en el caso bidimensional resulta habitual escribir la matriz de varianzas-covarianzas en función del coeficiente de correlación entre X e Y, ρ. Como $\sigma_{XY} = \rho \cdot \sigma_X \sigma_Y$, podemos escribir

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \sigma_Y \\ \rho \cdot \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} .$$

Según decíamos en la introducción a esta sección, las funciones de densidad de las normales bidimensionales se suelen exponer sin hacer uso de la notación matricial. Obsérvese que

$$|\Sigma| = \begin{vmatrix} \sigma_X^2 & \rho \cdot \sigma_X \sigma_Y \\ \rho \cdot \sigma_X \sigma_Y & \sigma_Y^2 \end{vmatrix} = \sigma_X^2 \sigma_Y^2 \cdot (1 - \rho^2)$$

y, por tanto,

$$\frac{1}{\sqrt{|\Sigma|}} = \frac{1}{\sigma_X \sigma_Y \sqrt{1 - \rho^2}} .$$

Además, la inversa de la matriz de varianzas-covarianzas es

$$\Sigma^{-1} = \frac{1}{\sigma_X^2 \sigma_Y^2 \cdot (1 - \rho^2)} \cdot \begin{pmatrix} \sigma_Y^2 & -\rho \cdot \sigma_X \sigma_Y \\ -\rho \cdot \sigma_X \sigma_Y & \sigma_X^2 \end{pmatrix},$$

con lo cual,

$$(x - \mu_X, y - \mu_Y)\Sigma^{-1}\begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix} =$$

$$= \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \rho^2)} (x - \mu_X, y - \mu_Y) \begin{pmatrix} \sigma_Y^2 & -\rho \cdot \sigma_X \sigma_Y \\ -\rho \cdot \sigma_X \sigma_Y & \sigma_X^2 \end{pmatrix} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix} =$$

$$= \frac{1}{1 - \rho^2} \left[\frac{(x - \mu_X)^2}{\sigma_X^2} - \frac{2\rho(x - \mu_X)(y - \mu_Y)}{\sigma_X \sigma_Y} + \frac{(y - \mu_Y)^2}{\sigma_Y^2} \right].$$

En definitiva, la función de densidad de una variable normal bidimensional es

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1 - \rho^2}} \cdot$$

$$\cdot \exp\left\{ -\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x - \mu_X}{\sigma_X} \right) \left(\frac{y - \mu_Y}{\sigma_Y} \right) + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

donde $-\infty < x < \infty$ y $-\infty < y < \infty$.

La representación gráfica de esta función de densidad bidimensional viene recogida en la figura 12.1.

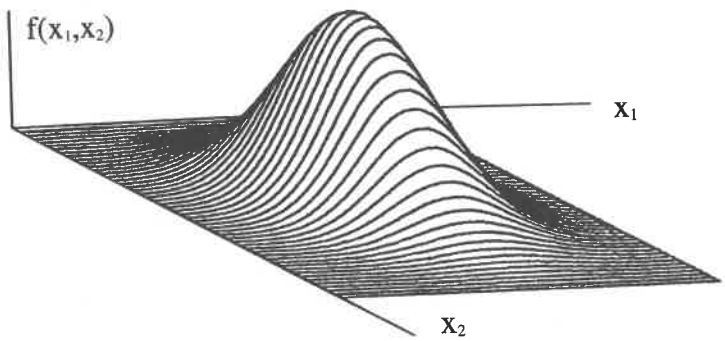


FIG. 12.1.

La proposición 12.6, referente a las transformaciones lineales de variables normales, puede concretarse para una normal bidimensional. Así, sea (X, Y) una normal bidimensional, $N_2(\mu, \Sigma)$, y sean U y V dos nuevas variables tales que

$$\begin{aligned} U &= aX + bY + c \\ V &= dX + eY + f \,. \end{aligned}$$

Esto se puede expresar matricialmente como

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} c \\ f \end{pmatrix}.$$

Aplicando la proposición 12.6, la variable bidimensional (U, V) seguirá una distribución normal bidimensional, con matriz de esperanzas $\mathbf{B}\mu + \mathbf{C}$ y matriz de varianzas-covarianzas $\mathbf{B}\Sigma\mathbf{B}'$, siempre que el rango de $\mathbf{B}$ sea igual a dos, donde

$$\mathbf{B} = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \quad \text{y} \quad \mathbf{C} = \begin{pmatrix} c \\ f \end{pmatrix}.$$

Obsérvese que la condición de rango acerca de $\mathbf{B}$ indica que $ae - bd$ tiene que ser distinto de cero, esto es, que U no es degenerada (es decir, a o b son no nulos), V no es degenerada (d o e no nulos) y, además, U y V no son combinaciones lineales una de la otra. Señalemos, también, que pueden escribirse explícitamente los elementos de la matriz de esperanzas y de la de varianzas-covarianzas de (U, V). En efecto, la primera resulta

$$\mathbf{B}\mu + \mathbf{C} = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \cdot \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} + \begin{pmatrix} c \\ f \end{pmatrix},$$

y, por tanto,

$$\mu_U = a\mu_X + b\mu_Y + c$$

y

$$\mu_V = d\mu_X + e\mu_Y + f \,.$$

La matriz de varianzas-covarianzas es

$$\mathbf{B}\Sigma\mathbf{B}' = \begin{pmatrix} a & b \\ d & e \end{pmatrix} \cdot \begin{pmatrix} \sigma_X^2 & \rho \cdot \sigma_X \sigma_Y \\ \rho \cdot \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ d & e \end{pmatrix}';$$

en consecuencia, operando en el producto matricial anterior tendremos que

$$\begin{aligned} \sigma_U^2 &= a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2 \,, \\ \sigma_V^2 &= d^2\sigma_X^2 + 2de\rho\sigma_X\sigma_Y + e^2\sigma_Y^2 \quad \text{y} \\ \sigma_{UV} &= ad\sigma_X^2 + (bd + ae)\rho\sigma_X\sigma_Y + be\sigma_Y^2 \,. \end{aligned}$$

Los momentos de (U, V) hubieran podido determinarse directamente aplicando las propiedades de las esperanzas, varianzas y covarianzas que obtuvimos en el capítulo 7. La proposición 12.6 aporta la afirmación de que la distribución conjunta de (U, V) es normal bidimensional, así como una mayor sencillez de las expresiones.

Hallemos, por último, la distribución de Y, condicionada por los valores de X, para una normal bidimensional (X,Y). Dicha distribución podría obtenerse concretando la proposición 12.10, pero resulta más ilustrativo su cálculo directo. Enunciemos y demostremos el resultado.

Proposición 12.11 *Sea (X,Y) una variable aleatoria con distribución normal bidimensional $N_2(\mu,\Sigma)$. La distribución de Y condicionada por $X=x$ es*

$$N_1\left(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),\sigma_Y\sqrt{1-\rho^2}\right)\,.$$

Demostración. La función de densidad condicionada es el cociente entre la conjunta y la marginal de X, es decir,

$$f_{(Y\mid X=x)}(y)=\frac{f_{(X,Y)}(x,y)}{f_X(x)}\,.$$

Efectuando dicho cociente, resulta, para $-\infty<x<\infty$,

$$f_{(Y\mid X=x)}(y)=\frac{1}{\sigma_Y\sqrt{1-\rho^2}\sqrt{2\pi}}\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right]^2\right\}=$$

$$=\frac{1}{\sigma_Y\sqrt{1-\rho^2}\sqrt{2\pi}}\cdot\exp\left\{\frac{-1}{2\sigma_Y^2(1-\rho^2)}\left[y-\left(\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}\cdot(x-\mu_X)\right)\right]^2\right\},$$

que es la función de densidad de una distribución normal unidimensional cuya media es $\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}\cdot(x-\mu_X)$ y cuya desviación típica es $\sigma_Y\sqrt{1-\rho^2}$. ♣

Comentábamos en este mismo capítulo que el hecho de que las dos distribuciones marginales fuesen normales no implicaba que la conjunta lo fuese. Ello se debe a que las marginales no caracterizan a la conjunta, salvo cuando se exige independencia entre las mismas. Ahora bien, conocida la distribución marginal de X y la distribución de Y condicionada por X, sí queda determinada de forma unívoca la distribución conjunta. Por tanto, *si X es una variable aleatoria con distribución normal $N(\mu_X,\sigma_X)$ y la distribución de Y condicionada por $X=x$ es, también, normal de media $\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}\cdot(x-\mu_X)$ y desviación típica $\sigma_Y\sqrt{1-\rho^2}$, la distribución conjunta de (X,Y) es una normal bidimensional $N_2(\mu,\Sigma)$, donde*

$$\mu=\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}\quad y\quad \Sigma=\begin{pmatrix}\sigma_X^2 & \sigma_{XY}\\ \sigma_{XY} & \sigma_Y^2\end{pmatrix}\,.$$

El siguiente corolario va a proporcionar un importante resultado sobre la regresión entre las componentes de una normal bidimensional.

Corolario 12.11.1 *Sea (X,Y) una variable aleatoria con distribución normal bidimensional $N_2(\mu,\Sigma)$. Entonces,*

$$E(Y\mid X=x)=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)=\mu_Y+\frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_X)\,,\qquad \textit{para } -\infty<x<\infty\,.$$

La demostración es inmediata, sin más que tener en cuenta la distribución condicionada, que es, según la proposición anterior, una normal, siendo su media $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\cdot(x-\mu_X)$.

Observe el lector que en este corolario hemos hallado la curva general de regresión de Y sobre X,

$$y = E(Y \mid X = x)\,,$$

resultando ser una recta, de donde se deduce que la curva y la recta de regresión coinciden. Entonces, podemos estudiar la bondad del ajuste por medio del cuadrado del coeficiente de correlación ρ, dado que, en estas circunstancias,

$$\rho^2 = \eta^2_{Y\mid X}\,.$$

Se deducirían análogos resultados si planteamos la distribución de X condicionada por Y.

Ejemplo 12.4. Hallemos, en la situación del ejemplo 12.1, el consumo esperado de una familia que tiene una renta mensual de 80 mil pesetas.

Siendo X_1 y X_2 la renta y el consumo mensuales, respectivamente. Sabemos que

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \hookrightarrow N_2\left[\begin{pmatrix} 100 \\ 90 \end{pmatrix}, \begin{pmatrix} 900 & 731.98 \\ 731.98 & 625 \end{pmatrix}\right].$$

Por el corolario anterior,

$$E(X_2 \mid X_1 = x_1) = 90 + \frac{731.98}{900}(x_1 - 100)\,.$$

Por tanto, para un valor de renta $x_1 = 80$, tendremos el siguiente consumo esperado:

$$E(Y \mid X = 80) = 90 + \frac{731.98}{900}(80 - 100) = 73.7 \text{ miles de pesetas.}$$

Si calculamos el coeficiente de correlación,

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{731.98}{\sqrt{900}\sqrt{625}} = 0.97\,,$$

observamos que existe una gran correlación positiva entre la renta y el consumo, lo cual, por otro lado, es perfectamente lógico. ◁

12.3 Ejercicios propuestos

12.1. Sea Σ una matriz definida positiva de orden n. Demuéstrese que existe una matriz A de orden n, no singular, tal que $\Sigma = AA'$. A dicha matriz A se le denomina *raíz cuadrada* de Σ. (Sugerencia: Descompóngase Σ como PDP', donde D es su correspondiente matriz diagonal, y escríbase D como producto de una matriz diagonal por sí misma.)

12.2. Sea $\mathbf{X}$ una variable aleatoria con distribución normal n-dimensional, $N_n(\mu, \Sigma)$, tal que sus componentes, $X_1, \ldots, X_n$, son independientes. Demuéstrese que la variable $(\mathbf{X}-\mu)'\Sigma^{-1}(\mathbf{X}-\mu)$ sigue una distribución χ^2 con n grados de libertad.

12.3. Sea (X_1, X_2, X_3) una variable aleatoria con distribución normal tridimensional de parámetros

$$\mu = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 1 & 2 \\ 2 & 2 & 9 \end{pmatrix} .$$

a) Sea una nueva variable tridimensional (Z_1, Z_2, Z_3) tal que

$$\begin{aligned} Z_1 &= X_1 - 2X_2 + 3X_3 \\ Z_2 &= 3X_1 - X_2 + 2X_3 \\ Z_3 &= -X_1 + 2X_2 + 2X_3 . \end{aligned}$$

¿Sigue (Z_1, Z_2, Z_3) una distribución normal tridimensional? Obténganse sus parámetros.

b) Sea una nueva variable

$$Z_4 = 3X_1 - X_2 + 7X_3 .$$

¿Es (Z_1, Z_2, Z_3, Z_4) una variable normal cuatridimensional? Calcúlese la esperanza y la matriz de varianzas-covarianzas de (Z_1, Z_2, Z_3, Z_4).

12.4. Sea

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \hookrightarrow N_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 9 \end{pmatrix} \right]$$

y sea

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} ,$$

donde

$$Y_1 = aX_1 + bX_2 \qquad \text{e} \qquad Y_2 = cX_1 + dX_2 .$$

¿Qué condiciones deben verificar a, b, c y d, para que $\mathbf{Y}'\mathbf{Y}$ se distribuya como una χ^2 con 2 grados de libertad?

12.5. Sea una variable aleatoria cuya función de densidad es

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{1}{\pi} e^{-\frac{1}{2}(x^2+y^2)} & \begin{cases} x < 0, y > 0 \\ x > 0, y < 0 \end{cases} \\ 0 & \text{en el resto} \end{cases}$$

a) Obténganse las distribuciones marginales de X e Y.

b) Calcúlese la covarianza entre X e Y.

12.6. Sean Y_1 e Y_2 variables aleatorias independientes con distribución $N(0,1)$. Sean las variables $X_1 = 3 + 2Y_1 - Y_2$ y $X_2 = 5 + Y_1 + Y_2$.

a) Hállese la función de densidad de (X_1, X_2).

b) ¿Cómo se distribuye la variable $Z = 3X_1 + 4X_2 - 1$?

12.7. El número de mujeres que mueren al año a causa de una cierta enfermedad es una variable aleatoria $N(\mu, 4)$, y el de hombres una $N(100, 5)$, siendo el coeficiente de correlación entre ambas de 0.8 (se supone que la distribución conjunta de las dos variables es una normal bidimensional).
Considerando un período de 10 años, y sabiendo que la probabilidad de que el número total de enfermos que mueren al año sea superior a 300 es de 0.5, calcúlese la probabilidad de que en 8 de esos 10 años, el número de mujeres que mueren sea superior a 150.

12.8. Las calificaciones del primer parcial de una asignatura se distribuyen como una normal de media 4 y desviación típica 1, y las del segundo parcial como una normal de media 6 y desviación típica 2. Suponemos que las notas de ambos parciales siguen una distribución normal bidimensional, siendo 0.8 el coeficiente de correlación.
Un alumno aprueba la asignatura si la media arimética de ambas calificaciones es al menos 5. Hállese la probabilidad de que en un grupo de 8 alumnos sólo haya 3 que aprueben la asignatura.

12.9. Un minorista vende por encargo previo de sus clientes dos productos A y B al precio de 35 y 60 pesetas. Dichos productos los compra en un almacén a 30 y 50 pesetas, respectivamente, gastando 100 pesetas en transporte.
Sabiendo que los pedidos diarios de A y B siguen distribuciones normales independientes de medias 10 y 20, y varianzas 9 y 4, respectivamente, hállese:

a) El coeficiente de correlación entre los gastos y los ingresos del minorista.

b) La probabilidad de que los ingresos dupliquen al menos a los gastos.

12.10. Una central eléctrica suministra energía a dos pueblos. El consumo doméstico (en miles de kilovatios) sigue una distribución normal de media 700 y desviación típica 30 para el primer pueblo, y de media 500 y desviación típica 20 para el segundo, siendo su distribución conjunta una normal bidimensional, con coeficiente de correlación lineal igual a 0.9.
Por otra parte, los consumos industriales son 1100 y 900 miles de kilovatios, para cada uno de los pueblos.

a) Si el consumo total entre ambos pueblos supera los 3300 miles de kilovatios, se producirá un apagón. ¿Cuál es la probabilidad de que esto ocurra?

b) El primer pueblo paga 0.3 unidades monetarias por cada mil kilovatios, cantidad que es 0.4 para el caso del segundo. Hállese la probabilidad de que la factura eléctrica del primer pueblo supere a la del segundo.

12.11. En un estudio realizado sobre la renta de las familias en las que ambos cónyuges trabajan, se ha estimado que el salario mensual (en miles de pesetas) del marido y de la mujer sigue una distribución normal bidimensional de medias 100 y 80 respectivamente, varianzas iguales y coeficiente de correlación 0.8. Sabemos que el 67 % de los hombres tiene un salario superior al de su pareja. Se toma una familia al azar y se consulta una de las nóminas, resultando ser inferior a 90 mil pesetas. ¿Cuál es la probabilidad de que corresponda al marido?

12.12. Una fábrica tiene 30 máquinas de tipo A y 20 de tipo B. Las máquinas de tipo A producen tornillos y arandelas cuyos pesos se distribuyen normalmente con medias 10 y 1 gramos y desviaciones típicas 0.5 y 0.1, respectivamente, siendo el coeficiente de correlación entre ambos pesos de 0.8. Las máquinas de tipo B producen tornillos y arandelas, siendo el peso total una variable aleatoria normal de media 11 gramos y desviación típica 0.6. Admitiendo que la distribución conjunta de los pesos de tornillos y arandelas es una variable aleatoria normal bidimensional.

a) ¿Cuál es la probabilidad de que la suma de ambos pesos sea mayor que 12 gramos?

b) Si consideramos sólo una máquina de tipo A, ¿cuál es la probabilidad de que el peso del tornillo sea al menos el doble que el de la arandela?

12.13. De un estudio realizado sobre el gasto telefónico mensual, en miles de pesetas, de dos familias, se han obtenido los siguientes resultados:
Familia 1: El gasto en llamadas urbanas es una distribución normal de media 5 y desviación típica 1, siendo el gasto en interurbanas de 7 mil pesetas.
Familia 2: El gasto en llamadas urbanas es una normal de media 8 y desviación típica 2, siendo el gasto en las interurbanas de 8 mil pesetas.
Suponiendo que los gastos urbanos siguen una distribución normal bidimensional, y considerando un período de un año, calcúlese la probabilidad de que en al menos dos meses los gastos telefónicos de la primera familia superen a los de la segunda, en cada una de las situaciones siguientes:

a) Cuando los gastos en llamadas urbanas de las dos familias son independientes.

b) Sabiendo que la probabilidad de que la suma de los gastos totales de ambas familias en un mes sea inferior a 32 mil pesetas, es igual a 0.9772.

12.14. Una empresa de electrodomésticos exporta a dos países dos tipos de frigoríficos, A y B.
En el país I obtiene 25 mil pesetas de ingreso por cada frigorífico de tipo A vendido, y 15 mil pesetas con los del tipo B, con unos gastos fijos mensuales de un millón de pesetas. En el país II consigue 20 mil de ingreso con cada frigorífico de cualquiera de los dos tipos, con gastos fijos de 550 mil pesetas.
El número de frigoríficos de cada tipo vendidos al mes sigue una distribución normal bidimensional de media 200 y varianza 100 para los del tipo A, y de media 100 y varianza 25 para los del tipo B, siendo el coeficiente de correlación -0.8.
Hállese la probabilidad de que el beneficio obtenido al mes sea mayor en el país I que en el país II.

12.15. Sea una variable aleatoria (X, Y) con distribución normal bidimensional cuyos parámetros μ y Σ son:

$$\mu = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}.$$

a) Hállese el coeficiente de correlación entre X y $X+Y$, y $p[X+Y \leq 5|X > 4]$.

b) Se definen las variables $U = 2X + Y$ y $V = X - Y$. Calcúlense las rectas de regresión de U y V.

12.16. Sea (X, Y) una variable aleatoria bidimensional. Supóngase que X sigue una distribución $N(0, 1)$ y que la distribución de Y condicionada por $X = x$ es normal de media $2x - 3$ y varianza 12. Obténgase la distribución marginal de Y y el coeficiente de correlación entre X e Y.

12.17. K. PEARSON estudió en 1903 la relación entre las alturas, en centímetros, de los padres, X, y de los hijos, Y, obteniendo que (X, Y) se ajustaba a una distribución normal bidimensional, siendo las medias de X e Y, 172 y 174.7, las desviaciones típicas, 6.91 y 6.98, y el coeficiente de correlación lineal $\rho = 0.51$. Si un padre mide 178 centímetros, ¿cuál es la altura esperada de su hijo? Coméntese la bondad de la predicción.

12.18. Sea f la función de densidad de una variable normal bidimensional $N(\mathbf{0}, \mathbf{I})$. Sea (X, Y) una variable aleatoria bidimensional cuya función de densidad g es:

$$g(x, y) = \begin{cases} kf(x, y) & 1 \leq x \leq 3, \quad 1 \leq y \leq 2 \\ 0 & \text{en el resto} \end{cases}$$

Calcúlese:

a) La función de densidad conjunta de X e Y.

b) Las funciones de densidad marginales de ambas.

c) La recta de regresión de X sobre Z, donde $Z = 2X - Y$.

12.19. Sea (X, Y) una variable normal bidimensional cuyo coeficiente de correlación es 0.4. Sean U y V las variables X e Y tipificadas.

a) Compruébese que los coeficientes de correlación entre X e Y y entre U y V coinciden.

b) Calcúlense las curvas generales de regresión entre U y V y coméntese la bondad de los ajustes.

12.20. Sea (X, Y) una variable aleatoria con función de densidad:

$$f_{(X,Y)}(x, y) = \begin{cases} kf_{(U,V)}(x, y) & -0.5 \leq x \leq 0.2\,, \quad 0.5 \leq y \leq 1 \\ 0 & \text{en el resto} \end{cases}$$

donde $f_{(U,V)}(x, y)$ es la función de densidad de una variable normal bidimensional de media $\mathbf{0}$ y matriz de varianzas-covarianzas I.

a) Calcúlese la función de densidad de (X, Y).

b) Hállense las funciones de densidad marginales de X y de Y.

c) Calcúlese la esperanza de V condicionada por $U + V = 2$.

Capítulo 13

LEYES DE LOS GRANDES NÚMEROS Y TEOREMA DEL LÍMITE CENTRAL

En el capítulo 10 hemos estudiado cómo se distribuía la media muestral,

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n},$$

de n variables aleatorias, $X_1, \ldots, X_n$, independientes e igualmente distribuidas, cuando éstas eran normales de esperanza μ y desviación típica σ. El resultado que entonces probábamos es que la distribución de $\overline{X}_n$ es $N(\mu, \frac{\sigma}{\sqrt{n}})$, esto es, la media muestral es, en este caso, normal con la misma esperanza que las variables X_i y cuya varianza disminuye proporcionalmente al número de variables.

Cuando n es suficientemente grande, la desviación típica de la media muestral es despreciable y, por tanto, $\overline{X}_n$ puede considerarse como una variable degenerada en el valor de μ.

Lo que este resultado pone de manifiesto es que, si se dispone de muchos agentes cuyos efectos individuales son aleatorios e independientes unos de otros, y se comportan cada uno de ellos de acuerdo con la misma distribución normal, el efecto promedio de todos ellos puede suponerse, en la práctica, conocido y no aleatorio. Este efecto promedio es, casi, el efecto esperado de uno cualquiera de los agentes.

El resultado anterior puede ilustrarse en términos numéricos. Si los efectos individuales (esto es, las variables X_i) se distribuyen, por ejemplo, como una $N(5, 1)$, la probabilidad de que uno de estos efectos esté entre 4.9 y 5.1 (esto es, suficientemente próximo a 5) es sólo de 0.07966; es decir, únicamente ocho de cada 100 efectos serán similares al efecto esperado, que vale cinco. Sin embargo, si disponemos, por ejemplo, de 1000 de estos efectos y son independientes, la probabilidad de que el efecto promedio se encuentre entre 4.9 y 5.1 es de 0.998, casi la unidad. El cuadro 13.1 ilustra los valores de dicha probabilidad para distinto número de efectos.

CUADRO 13.1

Número de efectos	Desviación típica del promedio	Esperanza del promedio	Probabilidad de que el promedio esté entre 4.9 y 5.1
1	1	5	0.07966
10	0.3162277	5	0.248174
100	0.1	5	0.68269
1000	0.0316227	5	0.9984342
5000	0.0141421	5	$1\text{-}1.5466\cdot10^{-12}$
10000	0.01	5	1*

* *STATGRAPHICS*©, paquete estadístico con el que se han calculado estas probabilidades, no es capaz de diferenciar el valor verdadero de la unidad.

La tabla mostrada en el cuadro 13.1 ilustra, por tanto, el hecho de que, para n grande, la media muestral de variables normales, independientes e idénticamente distribuidas, es sensiblemente igual a una constante o, con más precisión, que es *parecida* a una constante con probabilidad *alta.*

No obstante, no existe unanimidad de criterios entre los especialistas acerca de lo que significa el calificativo *alta* aplicado a una probabilidad. Si 1000 topógrafos calculan la distancia entre dos pueblos y concluyen que ésta (obtenida como promedio de las mediciones) es aproximadamente igual a cinco kilómetros con probabilidad 0.998, tal vez a la mayor parte de los vecinos esa probabilidad les parezca alta, y estén dispuestos a aceptar ese promedio como valor verdadero de la distancia. Sin embargo, si se sabe que cierta presa revienta cuando el agua embalsada rebasa los 5.1 hm^3, la garantía de que el volumen embalsado está entre 4.9 y 5.1 hm^3 con probabilidad 0.998, puede resultar insuficiente.

En definitiva, la afirmación de que la media muestral de variables normales es *prácticamente* igual a su esperanza cuando n es *grande*, a veces no es suficiente, y resulta más aconsejable reflejar el hecho de que su distribución es una $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Cuando las variables $X_1, \ldots, X_n$ no son normales, la media muestral tampoco lo es. Por ejemplo, supóngase que dichas variables miden la propensión de ciertos consumidores a adquirir un vehículo. Para ello, definimos

$$X_i : \begin{cases} 1 & \text{si el consumidor } i \text{ compra el vehículo} \\ 0 & \text{si no lo compra} \end{cases}$$

donde $i = 1, \ldots, n$.

Supongamos que los consumidores tienen el mismo comportamiento aleatorio y que sus actitudes son independientes (esto es, que las variables $X_1, \ldots, X_n$ son independientes e igualmente distribuidas). Supongamos, asimismo, que la propensión esperada, es decir,

$$E(X_i) = p[X_i = 1] ,$$

vale 0.3. Si se dispone de n consumidores, su media muestral,

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} ,$$

mide el promedio de vehículos adquiridos por los mismos. Obviamente, el número de vehículos que adquieren será $n\overline{X}_n = \sum_{i=1}^{n} X_i$.

Como $n\overline{X}_n$ es una variable binomial, $\mathcal{B}(n, 0.3)$, podríamos calcular, por ejemplo, la probabilidad de que, si el número de consumidores es 10, el número de vehículos comprados estuviese entre 2 y 4.

Ahora bien, probaremos en este capítulo que, si n es grande, ocurre un hecho similar al que describíamos en el comentario inicial para variables normales, esto es, que el promedio $\overline{X}_n$ es aproximadamente igual a la propensión esperada, 0.3.

¿Podremos evaluar, en este caso, la probabilidad de que el promedio de compras sea parecido a 0.3? O, concretando esta afirmación, ¿de que dicho promedio se encuentre entre 0.29 y 0.31? Si el número de consumidores es, por ejemplo, mil, esa probabilidad será la de que el número de vehículos comprados se encuentre entre 290 y 310, probabilidad que podríamos calcular directamente sin más que tener en cuenta que dicha variable tiene una distribución binomial $\mathcal{B}(1000, 0.3)$. Esos cálculos son, no obstante, tediosos, y, como demostraremos en este capítulo, para valores grandes de n se cumple aproximadamente el resultado que era exacto cuando las variables $X_1, \ldots, X_n$ eran normales. En concreto, la distribución de la media muestral $\overline{X}_n$ es aproximadamente normal, con esperanza μ y varianza $\frac{\sigma^2}{n}$, donde μ y σ^2 son, respectivamente, la esperanza y la varianza de X_i. En nuestro ejemplo, $\overline{X}_{1000}$ tendría, aproximadamente, una distribución $N(0.3, 0.01449)$ ya que, como $Var(X_i) = 0.3 \cdot 0.7 = 0.21$, entonces,

$$Var(\overline{X}_n) = \frac{0.21}{1000} = 0.00021 \,,$$

y, en consecuencia, su desviación típica sería $\sqrt{0.00021} = 0.01449$.

Por tanto, la probabilidad buscada podrá calcularse aproximadamente como

$$\begin{aligned} p[0.29 \leq \overline{X}_n \leq 0.31] &\approx p[0.29 \leq Y \leq 0.31] = \\ &= p[-0.6901 \leq Z \leq 0.6901] = 0.509872 \,, \end{aligned}$$

donde Y y Z son variables aleatorias con distribución normal, con $\mu_Y = 0.3$, $\sigma_Y = 0.01449$, $\mu_Z = 0$ y $\sigma_Z = 1$.

Para muchas otras distribuciones, la distribución exacta de la media muestral no es conocida, por lo que el recurso de la aproximación a la normal, si es posible, resulta inevitable.

En este capítulo estudiaremos primeramente las definiciones de convergencias de sucesiones de variables aleatorias, convergencias que suministran distintos enfoques sobre el comportamiento de dichas sucesiones para valores grandes de n.

A continuación, abordaremos el estudio de las medias muestrales. En primer lugar, veremos en qué condiciones estas medias pueden considerarse aproximadamente variables degeneradas, análisis que conducirá a las denominadas leyes de los grandes números. En segundo lugar, analizaremos cómo pueden hallarse aproximadamente probabilidades para medias muestrales de muchas variables, cuando su distribución exacta no es conocida o, aun siéndolo, no resulte operativa desde el punto de vista de los cálculos.

13.1 Convergencia de sucesiones de variables aleatorias

Definición 13.1 *Sea un espacio probabilístico* (Ω, α, p) *y sean las variables aleatorias* X_i $(i = 1, 2, \ldots)$,

$$X_i : \Omega \to I\!R\,,$$

definidas en dicho espacio.

Llamaremos sucesión de variables aleatorias a la sucesión de funciones

$$X_1, \ldots, X_n, \ldots\,,$$

que denotaremos por $\{X_n\}_{n=1}^{\infty}$.

Se definen, asimismo, sucesiones de variables n-dimensionales. Por ejemplo, una sucesión de variables bidimensionales es una sucesión de funciones, (X_n, Y_n), de Ω en $\mathbb{R}^2$, que sean variables aleatorias. Dicha sucesión la denotaremos como $\{(X_n, Y_n)\}_{n=1}^{\infty}$.

Consideremos la sucesión de variables $\{X_n\}_{n=1}^{\infty}$. Cada realización del experimento al que está asociado el espacio probabilístico (Ω, α, p), proporciona un resultado $\omega \in \Omega$, y para ese resultado podemos obtener los valores que toma la sucesión de variables; el resultado será una sucesión numérica, $\{x_n\}_{n=1}^{\infty}$, donde $x_n = X_n(\omega)$.

Cuando trabajemos con sucesiones de variables aleatorias, se supondrá conocida la función de distribución de la variable n-dimensional $(X_1, \ldots, X_n)$ para cualquier n; conoceremos, por tanto, $F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n)$ para todo n y para todo punto de $\mathbb{R}^n$, $(x_1, \ldots, x_n)$.

Veamos un ejemplo sencillo.

Ejemplo 13.1. Sea una urna con bolas blancas y negras en una cierta proporción; de ella realizamos extracciones sucesivas con reemplazamiento. Consideremos la variable

$$X_n : \left| \begin{array}{ll} 1 & \text{si la bola es blanca en la } n\text{-ésima extracción} \\ 0 & \text{si la bola es negra en la } n\text{-ésima extracción} \end{array} \right.$$

Entonces, $\{X_n\}_{n=1}^{\infty}$ es una sucesión de variables aleatorias.

Supongamos que realizamos esa serie de extracciones, obteniendo una bola blanca en la primera, negra en las dos siguientes, blanca en la cuarta, ... La sucesión de variables se concreta en la siguiente sucesión numérica:

$$1, 0, 0, 1, \ldots \triangleleft$$

Estudiemos, a continuación, los distintos tipos de convergencia en las sucesiones de variables aleatorias. Para ello, se desarrollan dos enfoques de naturaleza esencialmente distinta, adecuado cada uno de ellos a un cierto tipo de problemas. Por un lado, se estudia la convergencia de la sucesión de variables $\{X_n\}_{n=1}^{\infty}$ entendida como convergencia de funciones, es decir, considerando las variables aleatorias como funciones de Ω en $\mathbb{R}$. Por otro lado, se estudia la convergencia, no de las variables aleatorias, sino de las funciones de distribución, $F_n(x)$, asociadas a dichas

variables. El primer enfoque dará lugar a tres tipos de convergencia: casi segura, en probabilidad y en media de orden r; el segundo enfoque, a la convergencia en ley o en distribución. Aunque estos dos tipos de problemas son diferentes, tienen una relación entre ellos, como veremos más adelante. Estudiemos, pues, las distintas clases de convergencia.

13.1.1 Convergencia casi segura

Definición 13.2 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias definidas en un espacio probabilístico* (Ω, α, p), *y sea* X *una variable aleatoria sobre el mismo espacio. Se dice que* $\{X_n\}_{n=1}^{\infty}$ *converge casi seguro hacia* X *si*

$$p[\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)] = 1 .$$

Si $\{X_n\}_{n=1}^{\infty}$ converge casi seguro hacia X lo escribiremos como

$$X_n \xrightarrow{c.s.} X .$$

Este concepto *suaviza* el concepto de convergencia funcional, que exigiría la convergencia de $X_n(\omega)$ hacia $X(\omega)$, para todo $\omega \in \Omega$. En la convergencia casi segura puede haber puntos ω en los que $X_n(\omega)$ no converja hacia $X(\omega)$ (incluso en los que no converja hacia ningún otro punto), pero estos puntos forman un conjunto de probabilidad nula. Por tanto, el concepto de convergencia casi segura no difiere sustancialmente del concepto de convergencia funcional, sino que únicamente lo suaviza, probabilísticamente, en cierto sentido.

Seguidamente enunciaremos de forma sintética algunas propiedades de interés referentes a la convergencia casi segura, propiedades que no demostraremos:

P1. $X_n \xrightarrow{c.s.} X$ si, y sólo si, $X_n - X \xrightarrow{c.s.} 0$.

P2. Si para cualquier $\varepsilon > 0$, la serie

$$\sum_{n=1}^{\infty} p[|X_n - X| > \varepsilon]$$

es convergente, entonces,

$$X_n \xrightarrow{c.s.} X .$$

P3. Si $X_n \xrightarrow{c.s.} X$ y g es una función real de variable real continua, se cumple que

$$g(X_n) \xrightarrow{c.s.} g(X) .$$

En particular,

a) Si $X_n \xrightarrow{c.s.} X$, entonces $k \cdot X_n \xrightarrow{c.s.} k \cdot X$, para todo número real k.

b) Si $X_n \xrightarrow{c.s.} X$, entonces $X_n^k \xrightarrow{c.s.} X^k$, para todo número real k positivo.

P4. Si $X_n \xrightarrow{c.s.} X, Y_n \xrightarrow{c.s.} Y$ y g es una función continua de $\mathbb{R}^2$ en $\mathbb{R}$, entonces

$$g(X_n, Y_n) \xrightarrow{c.s.} g(X, Y) .$$

En particular,

a) Si $X_n \xrightarrow{c.s.} X$ e $Y_n \xrightarrow{c.s.} Y$, entonces $X_n + Y_n \xrightarrow{c.s.} X + Y$.

b) Si $X_n \xrightarrow{c.s.} X$ e $Y_n \xrightarrow{c.s.} Y$, entonces $X_n \cdot Y_n \xrightarrow{c.s.} X \cdot Y$.

c) Si $X_n \xrightarrow{c.s.} X$ e $Y_n \xrightarrow{c.s.} Y$, entonces $\frac{X_n}{Y_n} \xrightarrow{c.s.} \frac{X}{Y}$, siempre que los cocientes $\frac{X}{Y}$ y $\frac{X_n}{Y_n}$ estén definidos, esto es, siempre que los denominadores sean distintos de cero, salvo, a lo sumo, en un conjunto de probabilidad nula.

Ejemplo 13.2. Comprobemos, utilizando la segunda propiedad, que si

$$X_n : \left| \begin{array}{ll} 1 & p[X_n = 1] = \frac{1}{n^2} \\ 0 & p[X_n = 0] = 1 - \frac{1}{n^2} \end{array} \right.$$

la sucesión $\{X_n\}_{n=1}^{\infty}$ converge casi seguro hacia 0, es decir,

$$X_n \xrightarrow{c.s.} 0 .$$

Obviamente, para cualquier $\varepsilon > 0$,

$$p[X_n > \varepsilon] = \begin{cases} 0 & \text{si } \varepsilon \geq 1 \\ \frac{1}{n^2} & \text{si } 0 < \varepsilon < 1 \end{cases}$$

Por tanto,

$$\sum_{n=1}^{\infty} p[|X_n - 0| > \varepsilon] = \begin{cases} 0 & \text{si } \varepsilon \geq 1 \\ \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} & \text{si } 0 < \varepsilon < 1 \end{cases}$$

En cualquier caso, la serie es convergente y, consecuentemente, la sucesión $\{X_n\}_{n=1}^{\infty}$ converge casi seguro hacia la variable degenerada que siempre vale cero. ◁

13.1.2 Convergencia en probabilidad

La convergencia casi segura, anteriormente definida, resulta aún un concepto de convergencia muy exigente para lo que realmente necesitamos en muchas situaciones. Para paliar este inconveniente definimos una convergencia más *suave*.

Definición 13.3 *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y sea X una variable aleatoria, todas definidas sobre el mismo espacio probabilístico (Ω, α, p). Se dice que $\{X_n\}_{n=1}^{\infty}$ converge en probabilidad hacia X si, para cualquier $\varepsilon > 0$,*

$$\lim_{n \to \infty} p[|X_n - X| > \varepsilon] = 0 .$$

Dicho de otra forma, $\{X_n\}_{n=1}^{\infty}$ converge en probabilidad a X cuando, si ε es un número no negativo cualquiera,

$$\lim_{n\to\infty} p[|X_n - X| \leq \varepsilon] = 1 .$$

Utilizaremos la notación

$$X_n \xrightarrow{c.p.} X$$

para indicar que la sucesión $\{X_n\}_{n=1}^{\infty}$ converge en probabilidad a la variable aleatoria X. A veces, esta convergencia se simboliza, también, como

$$X_n \xrightarrow{p} X .$$

Enunciemos sucintamente algunas propiedades de la convergencia en probabilidad, que no demostraremos.

P1. $X_n \xrightarrow{c.p.} X$ si, y sólo si, $X_n - X \xrightarrow{c.p.} 0$.

P2. Si $X_n \xrightarrow{c.p.} X$ y g es una función real de variable real continua, entonces

$$g(X_n) \xrightarrow{c.p.} g(X) .$$

En particular,

a) Si $X_n \xrightarrow{c.p.} X$, entonces $k \cdot X_n \xrightarrow{c.p.} k \cdot X$, para todo número real k.

b) Si $X_n \xrightarrow{c.p.} X$, entonces $X_n^k \xrightarrow{c.p.} X^k$, para todo número positivo k.

P3. Si $X_n \xrightarrow{c.p.} X$, $Y_n \xrightarrow{c.p.} Y$ y g es una función continua de $\mathbb{R}^2$ en $\mathbb{R}$, entonces

$$g(X_n, Y_n) \xrightarrow{c.p.} g(X, Y) .$$

En particular,

a) Si $X_n \xrightarrow{c.p.} X$ e $Y_n \xrightarrow{c.p.} Y$, entonces $X_n + Y_n \xrightarrow{c.p.} X + Y$.

b) Si $X_n \xrightarrow{c.p.} X$ e $Y_n \xrightarrow{c.p.} Y$, entonces $X_n \cdot Y_n \xrightarrow{c.p.} X \cdot Y$.

c) Si $X_n \xrightarrow{c.p.} X$ e $Y_n \xrightarrow{c.p.} Y$, entonces $\frac{X_n}{Y_n} \xrightarrow{c.p.} \frac{X}{Y}$, siempre que los cocientes $\frac{X}{Y}$ y $\frac{X_n}{Y_n}$ estén definidos, esto es, siempre que los denominadores sean distintos de cero, salvo, a lo sumo, en un conjunto de probabilidad nula.

El siguiente resultado demuestra que ciertas sucesiones de variables aleatorias con la misma esperanza convergen en probabilidad a dicha esperanza.

Proposición 13.1 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias de media* μ_n *y varianza* σ_n^2, *tal que* $\lim_{n\to\infty} \sigma_n^2 = 0$. *Entonces,*

$$X_n - \mu_n \xrightarrow{c.p.} 0 .$$

Demostración. Como las variables tienen esperanza y varianza, podemos utilizar la desigualdad de Chebyshev. Así, para cualquier ε no negativo, se cumplirá que

$$p[|X_n - \mu_n| > \varepsilon] \leq \frac{\sigma_n^2}{\varepsilon^2} .$$

Si tomamos límites en la desigualdad anterior,

$$\lim_{n\to\infty} p[|X_n - \mu_n| > \varepsilon] \leq \frac{1}{\varepsilon^2} \cdot \lim_{n\to\infty} \sigma_n^2 = 0 .$$

En consecuencia,

$$\lim_{n\to\infty} p[|X_n - \mu_n| > \varepsilon] = 0 ,$$

esto es,

$$X_n - \mu_n \xrightarrow{c.p.} 0 . \clubsuit$$

Si las variables de la proposición anterior tienen una esperanza constante, es decir, si $\mu_n = \mu$, para todo n, entonces la proposición anterior garantiza que

$$X_n - \mu \xrightarrow{c.p.} 0$$

y, en consecuencia, utilizando la propiedad 1 anterior,

$$X_n \xrightarrow{c.p.} \mu .$$

Veamos los siguientes ejemplos, que ilustrarán sobre el uso de la convergencia en probabilidad.

Ejemplo 13.3. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias, cada una de ellas con función de densidad:

$$f_n(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & 0 < x < \theta \\ 0 & \text{en el resto} \end{cases}$$

Comprobemos que la sucesión converge en probabilidad hacia θ.

En efecto, sea ε un número real no negativo arbitrario, y calculemos la probabilidad

$$p[|X_n - \theta| \leq \varepsilon] = p[\theta - \varepsilon \leq X_n \leq \theta + \varepsilon] .$$

La función de distribución de X_n es

$$F_n(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^n}{\theta^n} & 0 < x < \theta \\ 1 & x \geq \theta \end{cases}$$

como se comprueba con facilidad, integrando la función de densidad. Por tanto, si $\varepsilon < \theta$,

$$p[|X_n - \theta| \leq \varepsilon] = F_n(\theta + \varepsilon) - F_n(\theta - \varepsilon) = 1 - \frac{(\theta - \varepsilon)^n}{\theta^n} .$$

Por consiguiente,

$$\lim_{n\to\infty} p[|X_n - \theta| \leq \varepsilon] = 1 - \lim_{n\to\infty} \frac{(\theta - \varepsilon)^n}{\theta^n} = 1 .$$

Evidentemente, si $\varepsilon \geq \theta$ también se verifica el resultado anterior. Por tanto,

$$X_n \xrightarrow{c.p.} \theta \, .$$

Nótese que podemos aplicar la proposición anterior. En efecto,

$$E(X_n) = \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{n+1} \, \theta \, .$$

Además,

$$E(X_n^2) = \int_0^\theta x^2 \cdot \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{n+2} \, \theta^2 \, ,$$

y, por tanto,

$$Var(X_n) = \frac{n}{(n+2)(n+1)^2} \, \theta^2 \, ,$$

Como

$$\lim_{n \to \infty} Var(X_n) = 0 \, ,$$

si aplicamos la proposición anterior,

$$X_n - \frac{n}{n+1} \theta \xrightarrow{c.p.} 0 \, .$$

Pero, como puede verse, la proposición 13.1 no permite comprobar la convergencia en probabilidad hacia θ. ◁

Ejemplo 13.4. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias tales que

$$X_n : \left| \begin{array}{ll} \frac{1}{n} & p\left[X_n = \frac{1}{n}\right] = \frac{n^2-1}{n^2} \\ n & p[X_n = n] = \frac{1}{n^2} \end{array} \right.$$

Veamos que esta sucesión de variables aleatorias converge en probabilidad hacia cero.

Dado un $\varepsilon > 0$ cualquiera, siempre se cumple que, si n es mayor que un cierto n_0, entonces $\frac{1}{n} < \varepsilon < n$, con lo cual el suceso $[|X_n| < \varepsilon]$ coincide con $[X_n = \frac{1}{n}]$, para valores de n mayores que n_0.

Así, para esos valores de n,

$$p[|X_n| < \varepsilon] = p\left[X_n = \frac{1}{n}\right] = \frac{n^2-1}{n^2} \, .$$

En consecuencia,

$$\lim_{n \to \infty} p[|X_n| < \varepsilon] = \lim_{n \to \infty} \left(\frac{n^2-1}{n^2} \right) = 1 \, ,$$

y, por tanto, la sucesión converge en probabilidad hacia cero. ◁

13.1.3 Convergencia en media de orden r

Definición 13.4 *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias y sea X una variable aleatoria, definidas todas sobre el mismo espacio probabilístico (Ω, α, p). Sea r un número entero positivo. Se dice que $\{X_n\}_{n=1}^{\infty}$ converge en media de orden r hacia X si*

$$\lim_{n \to \infty} E[|X_n - X|^r] = 0 \, ,$$

siempre que dicha esperanza exista.

Denotaremos esta convergencia como

$$X_n \xrightarrow{m.r} X \, .$$

Para $r = 1$, decimos que $\{X_n\}_{n=1}^{\infty}$ **converge en media** hacia X.

Para $r = 2$, que es el caso más importante, o al menos el que se estudia con una mayor frecuencia, decimos que $\{X_n\}_{n=1}^{\infty}$ **converge en media cuadrática** hacia X y lo denotaremos como

$$X_n \xrightarrow{m.c.} X \, ,$$

concretándose como

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0 \, .$$

Únicamente enunciamos el siguiente resultado, que no probaremos, para la convergencia en media de orden r.

Proposición 13.2 *Si* $X_n \xrightarrow{m.r} X$, *entonces* $X_n \xrightarrow{m.s} X$, *para cualquier* s *que sea menor que* r.

Ejemplo 13.5. Consideremos nuevamente la sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$ del ejemplo 13.2:

$$X_n : \left| \begin{array}{ll} 1 & p[X_n = 1] = \frac{1}{n^2} \\ 0 & p[X_n = 0] = 1 - \frac{1}{n^2} \end{array} \right.$$

Comprobemos que esta sucesión converge en media de cualquier orden hacia la variable aleatoria degenerada en el cero:

$$X_n \xrightarrow{m.r} 0 \, .$$

Efectivamente, como X_n es una variable no negativa,

$$\lim_{n\to\infty} E[|X_n - o|^r] = \lim_{n\to\infty} E[(X_n)^r] = \lim_{n\to\infty} \frac{1}{n^2} = 0 \, . \triangleleft$$

13.1.4 Convergencia en ley o en distribución

Si las anteriores convergencias eran de las variables, en el sentido de que comparaban los valores de las variables de la sucesión con los valores de la variable límite, la convergencia en ley, también llamada en distribución, es una convergencia de las distribuciones de probabilidad de las variables. Veamos la siguiente definición.

Definición 13.5 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias y sea* X *otra variable aleatoria. Sean* F_n *y* F *las funciones de distribución de* X_n *y de* X, *respectivamente. Se dice que* $\{X_n\}_{n=1}^{\infty}$ *converge en ley, o en distribución, hacia* X *si*

$$\lim_{n\to\infty} F_n(x) = F(x) \, ,$$

para todos los números reales x *en los que* F *sea continua.*

La convergencia en ley se denota como

$$X_n \xrightarrow{\mathcal{L}} X \, .$$

Si la sucesión $\{X_n\}_{n=1}^{\infty}$ converge en ley hacia la variable X, diremos que X_n se distribuye asintóticamente como la variable X, designándolo abreviadamente por

$$X_n \overset{asint.}{\hookrightarrow} X \, .$$

En muchas ocasiones, nos referiremos a dos tipos de distribuciones de las variables de una sucesión: la distribución exacta (distribución de una cualquiera de las variables que componen la sucesión, es decir, distribución de X_n para un n fijo) y **distribución asintótica** (distribución de la variable a la que la sucesión converge en ley). Como veremos más adelante, utilizaremos estas últimas distribuciones como distribución aproximada de X_n, para valores grandes de n.

Por abuso de lenguaje, si la distribución de X_n o de X tiene nombre propio, es decir, es una distribución tipo, en la notación abreviada de la convergencia en ley y de la distribución asintótica, sustituiremos, a veces, X_n, X, o ambas, por la denominación abreviada de su distribución.

Ejemplo 13.6. Estudiemos la convergencia en ley de la sucesión $\{X_n\}_{n=1}^{\infty}$

$$X_n : \left| \begin{array}{ll} 1 & p[X_n = 1] = \frac{1}{n^2} \\ 0 & p[X_n = 0] = 1 - \frac{1}{n^2} \end{array} \right.$$

sucesión que, como se ha visto en ejemplos anteriores, converge casi seguro y en media de orden r hacia cero (el lector puede probar que también converge en probabilidad).

La función de distribución de X_n es

$$F_n(x) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{n^2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Por su parte, la función de distribución de X, variable degenerada en cero, es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Entonces, para los valores de x entre 0 y 1,

$$\lim_{n \to \infty} F_n(x) = \lim_{n \to \infty} \left(1 - \frac{1}{n^2}\right) = 1 = F_X(x) \, ,$$

siendo obvio el resultado para el resto de valores. Por tanto,

$$X_n \xrightarrow{\mathcal{L}} 0 \, . \triangleleft$$

El siguiente teorema, que por su dificultad no demostraremos, permite comprobar la convergencia en ley, utilizando las funciones características.

Teorema 13.1 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias y sea* X *una variable aleatoria. Sean* φ_n *y* φ *las funciones características de* X_n *y* X*, respectivamente. Entonces, la sucesión* $\{X_n\}_{n=1}^{\infty}$ *converge en ley hacia* X *si, y sólo si,*

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t) ,$$

para cualquier número real t*.*

Obsérvese que la convergencia en ley de variables aleatorias es la de las funciones de distribución, y no la de las funciones de probabilidad o de densidad. Ello es así porque lo que se busca es la convergencia de las probabilidades, convergencia que queda garantizada por la de las funciones de distribución, pero no por la de las funciones de probabilidad o de densidad. Sin embargo, hay sucesiones de variables discretas para las que la convergencia de las funciones de probabilidad implica la de las de distribución y, en consecuencia, la convergencia en ley de las variables aleatorias. Veamos las siguientes proposiciones en este sentido.

Proposición 13.3 *La distribución binomial converge en ley a la distribución de Poisson, esto es*[1]*,*

$$\mathcal{B}(n,p) \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda)$$

cuando $n \to \infty$*,* $p \to 0$ *y* $n \cdot p = \lambda$*.*

Demostración. En la proposición 9.5 se demostró la convergencia de la función de probabilidad de una distribución binomial hacia la de una distribución de Poisson; esto es, si $\{X_n\}_{n=1}^{\infty}$ es una sucesión tal que $X_n \hookrightarrow \mathcal{B}(n,p)$ y $X \hookrightarrow \mathcal{P}(\lambda)$,

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} p[X_n = k] = p[X = k] .$$

Ahora, hemos de probar que, para $x \in \mathbb{R}$,

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} F_{X_n}(x) = F_X(x) .$$

En efecto, si k es el mayor entero menor o igual que x,

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} F_{X_n}(x) &= \lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} F_{X_n}(k) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} (p[X_n = 0] + \cdots + p[X_n = k]) = \\
&= \lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} p[X_n = 0] + \cdots + \lim_{\substack{n\to\infty \\ p\to 0 \\ np=\lambda}} p[X_n = k] = \\
&= p[X = 0] + \cdots + p[X = k] = F_X(k) = F_X(x) ,
\end{aligned}$$

1. Como ya habíamos adelantado, la siguiente expresión es un simbolismo para indicar que, si $\{X_n\}_{n=1}^{\infty}$ es una sucesión de variables aleatorias tales que X_n sigue una distribución $\mathcal{B}(n,p)$, y X es una variable con distribución $\mathcal{P}(\lambda)$, entonces,

$$X_n \xrightarrow{\mathcal{L}} X .$$

En los siguientes resultados emplearemos también este tipo de simbolismos.

es decir, se cumple la convergencia entre las funciones de distribución. En definitiva,

$$\mathcal{B}(n,p) \xrightarrow{\mathcal{L}} \mathcal{P}(\lambda) \ . \clubsuit$$

Esta proposición puede demostrarse, también, comprobando la convergencia de las funciones características, esto es, viendo que la función característica de una binomial $\mathcal{B}(n, \frac{\lambda}{n})$, tiende hacia la de una variable de Poisson, $\mathcal{P}(\lambda)$, cuando n tiende hacia ∞.

Proposición 13.4 *La distribución hipergeométrica converge en ley a la distribución binomial, esto es,*

$$\mathcal{H}(N,n,p) \xrightarrow{\mathcal{L}} \mathcal{B}(n,p) \ ,$$

cuando $N \to \infty$, $n \to \infty$ *y* $\frac{N_1}{N} \to p$.

En la proposición 9.11 demostramos la convergencia de sus funciones de probabilidad. Utilizando un razonamiento análogo al de la proposición anterior, se comprueba el resultado.

Otras convergencias en ley, a las que hemos aludido en los capítulos anteriores, son las siguientes, que enunciaremos simbólicamente.

1) Cuando $n \to \infty$,

$$\sqrt{2\chi_n^2} - \sqrt{2n-1} \xrightarrow{\mathcal{L}} N(0,1) \ .$$

2) Cuando $n_1 \to \infty$,

$$t_n \xrightarrow{\mathcal{L}} N(0,1) \ .$$

3) Cuando $n_2 \to \infty$,

$$F_{n_1,n_2} \xrightarrow{\mathcal{L}} \frac{\chi_{n_1}^2}{n_1} \ .$$

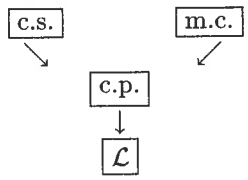
Asimismo, y sin pretender ser exhaustivos, existen relaciones de convergencia entre la distribución de Poisson y la distribución χ^2, entre la distribución geométrica y la distribución exponencial, y entre la binomial negativa y las distribuciones de Poisson y χ^2.

La proposición que a continuación enunciamos, y que tampoco probaremos, resume las relaciones entre las distintas convergencias que hemos definido.

Proposición 13.5 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias y sea* X *una variable aleatoria definida en un mismo espacio probabilístico.*

- *Si* $X_n \xrightarrow{c.s.} X$, *entonces* $X_n \xrightarrow{c.p.} X$.
- *Si* $X_n \xrightarrow{m.c.} X$, *entonces* $X_n \xrightarrow{c.p.} X$.
- *Si* $X_n \xrightarrow{c.p.} X$, *entonces* $X_n \xrightarrow{\mathcal{L}} X$.

El siguiente cuadro resume el enunciado de la proposición anterior.



Estas relaciones son las únicas existentes con carácter general. Ninguna de las convergencias implica otra que se encuentre más arriba en el cuadro anterior.

13.2 Leyes de los grandes números

Las leyes de los grandes números estudian el comportamiento probabilístico de las medias muestrales de variables aleatorias cuando el número de variables tiende hacia infinito. Más en concreto, si $\{X_n\}_{n=1}^{\infty}$ es una sucesión de variables aleatorias, se dice que esta sucesión cumple una ley de los grandes números si existe una sucesión de constantes, $\{a_n\}_{n=1}^{\infty}$, tal que la sucesión $\{\overline{X}_n - a_n\}_{n=1}^{\infty}$ converge hacia cero. Si la convergencia es casi segura, la ley se denomina fuerte, y se denomina débil cuando la convergencia es en probabilidad.

Las leyes de los grandes números tienen que ver con ciertas regularidades que se manifiestan cuando se realiza un gran número de veces el mismo experimento aleatorio. Hace ya varios siglos que se observó que las irregularidades que se presentaban en dichos experimentos tendían a compensarse si se promediaban todos sus efectos. Afirma G. V. GNEDENKO que «para muchos autores este fenómeno no parecía que exigiese ninguna explicación, ya que la presencia de regularidades, tanto en los fenómenos naturales como en los sociales, no era para ellos sino una manifestación de las reglas del orden divino»[2]. No obstante, los trabajos de CHEBYSHEV, MARKOV, KOLMOGOROV y otros, han descartado la universalidad del comportamiento regular de los promedios, orientando la investigación hacia el estudio de las condiciones en que se verifican las leyes de los grandes números.

13.2.1 Leyes débiles de los grandes números

Dijimos que una sucesión de variables aleatorias, $\{X_n\}_{n=1}^{\infty}$, verifica una ley débil de los grandes números, si existe una sucesión de constantes, $\{a_n\}_{n=1}^{\infty}$, tales que la sucesión $\{\overline{X}_n - a_n\}_{n=1}^{\infty}$ converge en probabilidad hacia cero. Usualmente, $a_n = \frac{1}{n}\sum_{i=1}^{n}\mu_i$, donde μ_i es la esperanza de X_i, que supondremos, siempre, que existe. Comencemos estudiando un resultado para variables independientes dos a dos, resultado que se conoce con el nombre de teorema de CHEBYSHEV.

2. G. V. GNEDENKO, *The Theory of Probability*, MIR, Moscú, 1969, p. 201.

Teorema 13.2 *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias independientes dos a dos, tales que sus varianzas existen y están acotadas por la misma constante. Entonces,*

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i \xrightarrow{c.p.} 0 ,$$

donde $\mu_i = E(X_i)$, $i = 1, 2, \ldots$

Demostración. Como las variables son independientes dos a dos,

$$Var(\overline{X}_n) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 ,$$

siendo $\sigma_i^2 = Var(X_i)$.

Si las varianzas están acotadas, es decir, $\sigma_i^2 \leq C$, para $i = 1, 2, \ldots$, entonces,

$$Var(\overline{X}_n) = \frac{1}{n^2}\cdot\sum_{i=1}^{n}\sigma_i^2 \leq \frac{1}{n^2}nC = \frac{C}{n} .$$

Por otro lado,

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n}E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu_i .$$

Aplicando la desigualdad de CHEBYSHEV,

$$p\left[\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > \varepsilon\right] \leq \frac{Var(\overline{X}_n)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} .$$

En consecuencia,

$$\lim_{n\to\infty} p\left[\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > \varepsilon\right] = 0 ,$$

y, por tanto, $\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i$ converge en probabilidad a cero. ♣

Obsérvese que en este teorema (y en el posterior corolario) bastaría exigir que las variables estuviesen incorrelacionadas; ahora bien, como las aplicaciones de éste y otros resultados se suelen hacer para variables independientes, lo enunciaremos de esta forma menos restrictiva.

El resultado que demostramos a continuación es un corolario del teorema de CHEBYSHEV que se utilizará con frecuencia en la Inferencia Estadística.

Corolario 13.2.1 *Si las variables son independientes dos a dos y tienen la misma esperanza,* μ*, y varianza,* σ^2*, entonces,*

$$\overline{X}_n \xrightarrow{c.p.} \mu .$$

Demostración. Como las varianzas son iguales, están acotadas. Además,

$$\frac{1}{n}\sum_{i=1}^{n}\mu_i = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu \, .$$

Aplicando el teorema de CHEBYSHEV,

$$\overline{X}_n - \mu \xrightarrow{c.p.} 0 \, ,$$

y, por tanto,

$$\overline{X}_n \xrightarrow{c.p.} \mu \, . \clubsuit$$

De hecho, si las variables están igualmente distribuidas y son independientes dos a dos, se cumplen las condiciones exigidas en el corolario y, consecuentemente,

$$\overline{X}_n \xrightarrow{c.p.} \mu \, .$$

El matemático ruso KHINCHIN demostró que, en este caso (independencia dos a dos e igual distribución de las variables), no es necesario exigir la existencia de las varianzas. Enunciemos el resultado, que se conoce como teorema de KHINCHIN, aunque no lo probaremos.

Teorema 13.3 *Si las variables de la sucesión* $\{X_n\}_{n=1}^{\infty}$ *son independientes dos a dos e igualmente distribuidas con esperanza* μ*, entonces,*

$$\overline{X}_n \xrightarrow{c.p.} \mu \, .$$

Ejemplo 13.7. Supongamos que las variables de $\{X_n\}_{n=1}^{\infty}$ recogen el resultado de una sucesión de tiradas de Bernoulli independientes, siendo p la probabilidad de éxito en cada tirada. Entonces,

$$\overline{X}_n \xrightarrow{c.p.} p \, .$$

Este ejemplo, denominado teorema de BERNOULLI, es un caso particular del teorema de CHEBYSHEV (así como del de KHINCHIN). Las variables son independientes (por tanto, lo son dos a dos), su esperanza $E(X_i) = p$ y su varianza, $Var(X_i) = p \cdot q$, existen; además, la esperanza y la varianza son las mismas para todas las variables de la sucesión. Aplicando el corolario del teorema de CHEBYSHEV,

$$\overline{X}_n \xrightarrow{c.p.} p \, .$$

Asimismo, como las variables son independientes dos a dos e igualmente distribuidas, siendo su esperanza p, podemos obtener el mismo resultado aplicando el teorema de KHINCHIN. $\triangleleft$

El ejemplo anterior ilustra el alcance de la denominada ley de regularidad estadística. En el capítulo 2 exigíamos a un experimento, para que fuese aleatorio, que cumpliera dicha ley, que describíamos intuitivamente como un comportamiento estable *a largo plazo*, esto es, cuando el experimento se desarrolla en las mismas condiciones muchas veces. Supongamos que se quiere estudiar la estabilidad a largo

plazo de un suceso, A, ligado a un experimento aleatorio. Nos planteamos realizar repetidamente, y de forma independiente, muchas veces el mismo experimento, y observar la frecuencia con la que se verifica el suceso A, es decir, el cociente entre el número de veces que ocurre A y el número de realizaciones del experimento, designando a esta frecuencia como $fr(A)$.

Si construimos una variable asociada a cada realización del experimento, X_i, de forma que $X_i = 1$, si en el experimento i-ésimo ocurre el suceso A, y $X_i = 0$, en caso de que no ocurra, dichas variables son independientes y con la misma distribución; de hecho, son variables de Bernoulli independientes, con esperanza igual a la probabilidad del suceso A. Además, la frecuencia $fr(A)$ coincide con la media muestral $\overline{X}_n$:

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\text{Número de veces que ocurre } A}{\text{Número de experimentos}} = fr(A) \, .$$

Entonces, el ejemplo anterior indica que

$$fr(A) = \overline{X}_n \xrightarrow{c.p.} p(A) \, ,$$

esto es, que, con probabilidad alta (siempre que n sea grande), la frecuencia del suceso A estará próxima a la probabilidad del mismo.

Aunque la demostración dada por CHEBYSHEV de su teorema es de 1866, el resultado que propone este ejemplo fue demostrado por J. BERNOULLI a finales del siglo XVII, siendo publicado después de su muerte (en 1713) en su libro *Ars Conjectandi*. Su interés residía en explicar los motivos por los que la frecuencia de un suceso se estabilizaba cuando el experimento se repetía muchas veces. Esta estabilidad se había observado en los juegos de azar; es más, según relata GNEDENKO, estudiosos del Cálculo de Probabilidades habían dedicado tiempo a observar este fenómeno[3].

Nosotros, probablemente más perezosos, no hemos dedicado el mismo tiempo a lanzar una moneda. Sin embargo, hemos simulado en el ordenador 5000 lanzamientos independientes de una moneda perfecta (con probabilidad de cara igual a 0.5). El teorema de BERNOULLI indica que la probabilidad de que la media muestral (esto es, la frecuencia del resultado *obtener cara*) se separe de 0.5 más de una cantidad ε (arbitraria, pero fija) tiende hacia cero cuando n tiende a infinito. Es decir, para cualquier ε positivo,

$$\lim_{n\to\infty} p[|\overline{X}_n - p| > \varepsilon] = \lim_{n\to\infty} p[|fr(A) - 0.5| > \varepsilon] = 0 \, ,$$

3. En el siglo XVIII, el naturalista francés BUFFON lanzó una moneda 4040 veces y obtuvo cara en 2048 ocasiones. En el experimento de BUFFON, la frecuencia del suceso *obtener cara* es, aproximadamente, de 0.507.

El estadístico inglés K. PEARSON lanzó una moneda 12000 veces y obtuvo 6019 caras. La frecuencia del suceso *obtener cara* en el experimento de Pearson es 0.5016. En otra ocasión, lanzó una moneda 24000 veces y obtuvo cara en 12012 ocasiones, resultando una frecuencia del suceso *obtener cara* de 0.5005.

En todos esos experimentos, la frecuencia del suceso se separó sólo levemente de su probabilidad, 0.5. (Tomado del libro de B. V. GNEDENKO, *op. cit.*, p. 205.)

CUADRO 13.2. *Resultados del lanzamiento simulado de una moneda*

Número de tirada	Número de caras	Frecuencia (en %)	Número de tirada	Número de caras	Frecuencia (en %)
1	1	100.00	28	18	64.29
2	1	50.00	29	19	65.52
3	2	66.67	30	20	66.67
4	2	50.00	40	24	60.00
5	2	40.00	50	29	58.00
6	2	33.33	60	32	53.33
7	3	42.86	70	38	54.29
8	3	37.50	80	44	55.00
9	3	33.33	90	52	57.78
10	3	30.00	100	55	55.00
11	4	36.36	150	78	52.00
12	4	33.33	200	106	53.00
13	5	38.46	250	136	54.40
14	6	42.86	300	156	52.00
15	7	46.67	350	179	51.14
16	8	50.00	400	201	50.25
17	9	52.94	450	227	50.44
18	10	55.56	500	255	51.00
19	10	52.63	600	297	49.50
20	11	55.00	700	343	49.00
21	12	57.14	800	390	48.75
22	13	59.09	900	439	48.78
23	14	60.87	1000	496	49.60
24	15	62.50	2000	997	49.85
25	16	64.00	3000	1469	48.97
26	17	65.38	4000	1979	49.48
27	17	62.96	5000	2490	49.80

donde A denota el suceso *obtener cara*. En definitiva, aunque la tendencia a cero de la probabilidad no tiene por qué ser monótona, para $n = 5000$ debemos esperar que dicha probabilidad sea tan pequeña que difícilmente dejará de observarse la aproximación en nuestro experimento simulado. El cuadro 13.2 resume los resultados obtenidos.

Debe distinguirse lo que es altamente probable de lo necesario. En este sentido, obsérvese que el comentario anterior no contradice la posibilidad de que 5000 lanzamientos de la moneda concluyesen con la obtención de otras tantas caras; ahora bien, esta posibilidad es muy improbable.

13.2.2 Convergencia en probabilidad de los momentos muestrales

En el estudio de la Inferencia Estadística, nos encontraremos con sucesiones de variables independientes e igualmente distribuidas. Una buena parte de las técnicas de estimación y contraste que allí se abordan, utilizan ciertas variables aleatorias que se obtienen a partir de la sucesión, y que reciben el nombre de características muestrales. A continuación definiremos las más usuales.

1) Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias; definimos:

- **Momento muestral de orden k centrado respecto del origen (o no central)**,

$$A_k = \frac{\sum_{i=1}^n X_i^k}{n} .$$

Para $k = 1$, dicho momento no es otro que la media muestral, $\overline{X}_n$.

- **Momento muestral de orden k centrado respecto de la media (o central)**,

$$M_k = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^k}{n} .$$

Para $k = 2$, este momento recibe el nombre de varianza muestral,

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n} ,$$

siendo su raíz cuadrada positiva, S_X, la desviación típica muestral.

2) Sea $\{(X_n, Y_n)\}_{n=1}^{\infty}$ una sucesión de variables aleatorias bidimensionales; definimos:

- **Covarianza muestral**,

$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{n} .$$

- **Coeficiente de correlación muestral**,

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} ,$$

donde S_X y S_Y son las desviaciones típicas muestrales de X y de Y, respectivamente.

Obsérvese que todas estas características muestrales son variables aleatorias que, cuando se realicen los experimentos que dan lugar a las sucesiones, tomarán valores concretos. De hecho, las características muestrales son, ellas mismas, sucesiones de variables aleatorias.

La siguiente proposición muestra que, si las variables aleatorias de una sucesión son independientes y están igualmente distribuidas, los momentos muestrales convergen, bajo ciertas condiciones, hacia los denominados momentos poblacionales, definidos éstos como los momentos de las variables de la sucesión. Como consecuencia, las características muestrales, función de dichos momentos, convergerán, a su vez, hacia las correspondientes características poblacionales.

Teorema 13.4 *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes dos a dos e igualmente distribuidas, y supongamos que existe el momento de orden k, $\alpha_k = E(X_n^k)$. Entonces,*

$$A_k \xrightarrow{c.p.} \alpha_k$$

y

$$M_k \xrightarrow{c.p.} \mu_k ,$$

donde $\mu_k = E[(X_n - E(X_n))^k]$.

Asimismo, si para una sucesión de variables bidimensionales, $\{(X_n, Y_n)\}_{n=1}^{\infty}$, *independientes dos a dos e igualmente distribuidas, existen* $\alpha_{2,0} = E(X_n^2)$, $\alpha_{0,2} = E(Y_n^2)$ *y* $\alpha_{1,1} = E(X_n \cdot Y_n)$, *entonces,*

$$S_{XY} \xrightarrow{c.p.} \sigma_{XY}$$

y

$$R_{XY} \xrightarrow{c.p.} \rho_{XY} ,$$

donde σ_{XY} *y* ρ_{XY} *son, respectivamente, la covarianza y el coeficiente de correlación lineal de* (X_n, Y_n).

Demostración. Probaremos que

$$A_k \xrightarrow{c.p.} \alpha_k .$$

Sea $Z_n = X_n^k$; entonces,

$$A_k = \frac{\sum_{i=1}^n X_i^k}{n} = \frac{\sum_{i=1}^n Z_i}{n} = \overline{Z}_n .$$

Podemos aplicar el teorema de KHINCHIN a las variables Z_n y, por tanto,

$$A_k = \overline{Z}_n \xrightarrow{c.p.} E(Z_n) = E(X_n^k) = \alpha_k .$$

El resto de los momentos y características muestrales se pueden poner como función (sumas, restas, productos, cocientes, ...) de momentos centrados respecto del origen y, consecuentemente, teniendo en cuenta las propiedades de la convergencia en probabilidad, dichos momentos y características muestrales convergen en probabilidad hacia sus correspondientes momentos y características poblacionales. Por ejemplo, la varianza muestral converge en probabilidad hacia la varianza poblacional. Así,

$$M_2 = S_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - (\overline{X})^2 = A_2 - (\overline{X})^2 .$$

Como

$$A_2 \xrightarrow{c.p.} \alpha_2$$

y

$$\overline{X} \xrightarrow{c.p.} \mu ,$$

basta aplicar las propiedades 2 y 3 de la convergencia en probabilidad para afirmar que

$$A_2 - (\overline{X})^2 \xrightarrow{c.p.} \alpha_2 - \mu^2 ,$$

esto es,

$$S_X^2 \xrightarrow{c.p.} \sigma_X^2 ,$$

donde σ_X^2 es la varianza de las variables de la sucesión. ♣

Estas convergencias habrían podido probarse, asimismo, para variables independientes dos a dos y con los mismos momentos para cualquier n (no necesariamente con la misma distribución), siempre que existiera el momento de orden doble al del que queremos buscar la convergencia. Bastaría, para ello, utilizar el corolario del teorema de CHEBYSHEV, en lugar del teorema de KHINCHIN.

El lector puede intuir la definición de otras características muestrales (coeficientes de asimetría y de apuntamiento, coeficiente de variación, ...) y su convergencia en probabilidad hacia el coeficiente poblacional correspondiente.

Todos estos resultados conducen a una idea: las características muestrales se aproximan a las características poblacionales, cuando el número de variables es grande, o mejor dicho, las características muestrales toman, en esta situación, valores cercanos a las características poblacionales con alta probabilidad. Este hecho es uno de los pilares básicos en el que se sustenta la Inferencia Estadística.

13.2.3 Leyes fuertes de los grandes números

En 1909, el matemático francés E. BOREL demostró una generalización del teorema de BERNOULLI. Como acabamos de indicar, este teorema establece que la frecuencia de éxito en n tiradas de Bernoulli independientes *se parece* a la probabilidad de éxito, p, con una probabilidad que tiende a la unidad, esto es, que dicha frecuencia converge en probabilidad a p. El resultado que demostró BOREL indica que dicha frecuencia también converge casi seguro hacia p.

Recordemos que la convergencia casi segura es un tipo de convergencia puntual, más fuerte, por tanto, que la convergencia en probabilidad. Para variables de Bernoulli independientes, BOREL probó que

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i \xrightarrow{c.s.} 0\ ,$$

donde, como antes, $\mu_i = E(X_i)$, $i = 1, 2, \ldots$

La demostración por BOREL de su teorema empujó a otros estadísticos a buscar condiciones más generales en las que se cumpliera el resultado anterior, es decir, condiciones para las cuales una sucesión verifica la denominada ley fuerte de los grandes números.

Los resultados en este sentido se deben esencialmente a KOLMOGOROV y no los probaremos en este libro. Enunciaremos, en primer lugar, el teorema de BOREL y, más adelante, haremos lo mismo con los teoremas de KOLMOGOROV.

Teorema 13.5 *Si las variables de la sucesión $\{X_n\}_{n=1}^{\infty}$ son variables aleatorias de Bernoulli, independientes y con el mismo parámetro p, entonces cumplen la ley fuerte de los grandes números y, por tanto,*

$$\overline{X}_n \xrightarrow{c.s.} p\ .$$

Demostración. Obsérvese que, como vimos en el teorema de BERNOULLI al estudiar las leyes débiles,

$$\frac{1}{n}\sum_{i=1}^{n}\mu_i = \frac{1}{n}np = p\ ,$$

y, por consiguiente, la convergencia que garantiza la ley fuerte,

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i = \overline{X}_n - p \xrightarrow{c.s.} 0\ ,$$

es equivalente a

$$\overline{X}_n \xrightarrow{c.s.} p \,. \clubsuit$$

Enunciemos, a continuación, el teorema de KOLMOGOROV, válido para variables independientes cuyas varianzas sean de un *tamaño razonable.*

Teorema 13.6 *Si las variables de la sucesión* $\{X_n\}_{n=1}^{\infty}$ *son independientes, con varianza* σ_n^2 *y, además, la serie*

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2}$$

es convergente, la sucesión cumple la ley fuerte de los grandes números, esto es,

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n} \mu_i \xrightarrow{c.s.} 0 \,.$$

Del teorema de Kolmogorov se deduce inmediatamente el siguiente corolario.

Corolario 13.6.1 *Si las variables de la sucesión* $\{X_n\}_{n=1}^{\infty}$ *son independientes y sus varianzas están acotadas por la misma constante,* C*, entonces se cumple la ley fuerte de los grandes números.*

Demostración. En este caso,

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \leq \sum_{n=1}^{\infty} \frac{C}{n^2} = C \sum_{n=1}^{\infty} \frac{1}{n^2} \,.$$

Pero, como es conocido, la última serie es convergente, y por consiguiente lo será, también, la primera. Aplicando el teorema de KOLMOGOROV, se cumplirá la ley fuerte de los grandes números, esto es,

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^{n} \mu_i \xrightarrow{c.s.} 0 \,. \clubsuit$$

Comparemos este resultado con el teorema de CHEBYSHEV que vimos en la subsección 13.2.1. Dicho teorema garantizaba que, si las varianzas estaban acotadas, se cumplía la ley débil de los grandes números para variables independientes dos a dos; el corolario que acabamos de probar garantiza no sólo la convergencia en probabilidad, sino la convergencia casi segura, pero exige que las variables sean mutuamente independientes.

Obviamente, si las variables están igualmente distribuidas y poseen varianza, ésta es la misma para todas ellas y, por tanto, se verifica la condición que exigíamos en el corolario anterior. No obstante, KOLMOGOROV demostró, también, que, si las variables tienen todas la misma distribución, no hace falta exigir condición alguna sobre las varianzas; más aún, no hace falta, ni siquiera, exigir que las varianzas existan. Enunciemos, sin demostración, el resultado.

Teorema 13.7 *Si las variables de la sucesión* $\{X_n\}_{n=1}^{\infty}$ *son independientes y están igualmente distribuidas, la existencia de la esperanza,* μ*, es una condición necesaria y suficiente para que se cumpla la ley fuerte de los grandes números, esto es, para que*

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{c.s.} \mu \, .$$

13.3 Teorema del límite central

En la introducción a este capítulo vimos cómo podía calcularse la probabilidad de que la media muestral (el promedio) de n variables aleatorias normales, independientes e igualmente distribuidas, se distanciase de la esperanza en menos de una cantidad. Ello era posible porque, en ese caso, la media muestral era una variable normal y, por tanto, su tipificación suministraba una variable normal estándar.

Más adelante, sugerimos que si las variables de la sucesión $\{X_n\}_{n=1}^{\infty}$ eran de Bernoulli, este resultado iba a ser aproximadamente cierto, en el sentido de que la distribución de la media muestral tipificada era aproximadamente la de una $N(0,1)$. Estas aproximaciones, que lo son de la función de distribución, se obtienen cuando se demuestran ciertas convergencias en ley de las medias muestrales tipificadas hacia la distribución normal estándar.

Las condiciones en que esto ocurre para una sucesión de variables aleatorias son diversas, y proporcionan los denominados *teoremas del límite central.* En concreto, dadas $X_1, \ldots, X_n, \ldots$ variables aleatorias con media y varianza, se dice que la sucesión $\{X_n\}_{n=1}^{\infty}$ verifica el teorema del límite central si

$$\frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{Var(\sum_{i=1}^{n} X_i)}} \xrightarrow{\mathcal{L}} N(0,1) \, .$$

Las diversas versiones del teorema varían en las hipótesis exigidas a las variables aleatorias (dependencia o independencia, distinta distribución o igual distribución, etcétera).

La primera versión del teorema, sólo para variables de Bernoulli con probabilidad de éxito $p = \frac{1}{2}$, fue demostrada de forma rudimentaria por A. DE MOIVRE a comienzos del siglo XVIII (en 1733), y generalizada por LAPLACE[4] para cualquier probabilidad p de éxito entre 0 y 1; este resultado se conoce como teorema de MOIVRE-LAPLACE. En 1901, LYAPUNOFF da la primera demostración rigurosa

4. Lo que de hecho demostraron estos autores fue que, si X es una variable aleatoria con distribución binomial $\mathcal{B}(n,p)$,

$$\lim_{n\to\infty} \frac{p[X=m]}{\frac{1}{\sqrt{2\pi npq}} \cdot e^{-\frac{1}{2}\frac{(m-np)^2}{npq}}} = 1 \, ,$$

siendo la convergencia uniforme si $\frac{m-np}{\sqrt{npq}}$ pertenece a un intervalo finito.

El lector reconocerá en el denominador de esta expresión la función de densidad de la variable normal de media np y varianza npq. A partir de este resultado se prueba la convergencia en distribución, aunque es más fácil probarlo con otras herramientas, como es la función característica.

de dicho resultado. Veinte años después, LINDEBERG y LEVY demuestran un teorema del límite central para variables independientes e igualmente distribuidas. LYAPUNOV probó, más tarde, una versión para variables independientes cualesquiera, siempre que cumplan una condición denominada, por motivos obvios, de LYAPUNOV.

A lo largo del presente siglo sigue habiendo diversas aportaciones. Nosotros demostraremos en este libro la versión de LEVY y LINDEBERG, que es suficiente para muchas de las aplicaciones.

Teorema 13.8 *Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes e igualmente distribuidas con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$. Entonces,*

$$\frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{Var(\sum_{i=1}^n X_i)}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma \cdot \sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1) .$$

Demostración. Llamemos Z_n a la variable

$$Z_n = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{Var(\sum_{i=1}^n X_i)}} ,$$

es decir,

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma \cdot \sqrt{n}} .$$

En virtud del teorema 13.1, basta con demostrar que

$$\lim_{n\to\infty} \varphi_{Z_n}(t) = \varphi_Z(t) ,$$

siendo φ_Z la función característica de la normal estándar.

Ahora bien, Z_n puede expresarse como

$$Z_n = \sum_{i=1}^n \frac{X_i - \mu}{\sigma \cdot \sqrt{n}} ,$$

y utilizando conocidas propiedades de las funciones características, se tiene:

$$\varphi_{Z_n}(t) = \varphi_{\sum_{i=1}^n \frac{X_i-\mu}{\sigma\cdot\sqrt{n}}}(t) = \left[\varphi_{\frac{X-\mu}{\sigma\cdot\sqrt{n}}}(t)\right]^n = \left[\varphi_{X-\mu}(\frac{t}{\sigma\sqrt{n}})\right]^n = \left[\varphi_Y(\frac{t}{\sigma\sqrt{n}})\right]^n ,$$

donde X es una cualquiera de las variables de la sucesión e $Y = X - \mu$ es una variable de esperanza 0 y varianza σ^2.

En la proposición 8.5 señalamos que, si existía el momento no centrado de orden k para una variable aleatoria X, su función característica podía escribirse como

$$\varphi_X(t) = \sum_{j=0}^k \frac{(it)^j}{j!}\alpha_j + o(t^k) ,$$

para todo número real t, y donde $\lim_{t\to 0} \frac{o(t^k)}{t^k} = 0$.

Podemos aplicar el resultado anterior a la variable Y, de la que conocemos la existencia de sus dos primeros momentos; en concreto, para $k = 2$,

$$\begin{aligned}\varphi_Y(t) &= \frac{(it)^0}{0!}\alpha_0 + \frac{(it)^1}{1!}\alpha_1 + \frac{(it)^2}{2!}\alpha_2 + o(t^2) = 1 + itE(Y) - \frac{t^2}{2}E(Y^2) + o(t^2) = \\ &= 1 - \frac{t^2\sigma^2}{2} + o(t^2)\,,\end{aligned}$$

donde $o(t^2)$ es una cantidad que tiende a cero más rápidamente que t^2. Por tanto,

$$\begin{aligned}\varphi_{Z_n}(t) = \left[\varphi_Y\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n &= \left[1 - \frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^2\sigma^2}{2} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]^n = \\ &= \left[1 - \frac{t^2}{2n} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]^n\,,\end{aligned}$$

y, tomando límites,

$$\lim_{n\to\infty}\varphi_{Z_n}(t) = \lim_{n\to\infty}\left[1 - \frac{t^2}{2n} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]^n\,.$$

Dado que se trata del límite de una potencia en la que la base tiende hacia uno y el exponente hacia infinito, resolvámoslo aplicando que

$$A^B = e^{B\ln(A)}\,.$$

Así,

$$\lim_{n\to\infty}\varphi_{Z_n}(t) = \exp\left\{\lim_{n\to\infty} n\cdot\ln\left[1 - \frac{t^2}{2n} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]\right\}\,,$$

lo que, teniendo en cuenta un resultado habitual en el cálculo de límites (cuando x tiende hacia cero, $\ln(1+x)$ y x son infinitésimos equivalentes), se convierte en

$$\lim_{n\to\infty}\varphi_{Z_n}(t) = \exp\left\{\lim_{n\to\infty} n\cdot\left[-\frac{t^2}{2n} + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]\right\}\,.$$

Es decir,

$$\lim_{n\to\infty}\varphi_{Z_n}(t) = \exp\left\{\lim_{n\to\infty}\left[-\frac{t^2}{2} + n\cdot o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]\right\} = \exp\left\{-\frac{t^2}{2}\right\}\,,$$

ya que

$$\lim_{n\to\infty} n\cdot o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right) = \lim_{n\to\infty} n\cdot\left(\frac{t}{\sigma\sqrt{n}}\right)^2\cdot\frac{o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)}{\left(\frac{t}{\sigma\sqrt{n}}\right)^2} = \frac{t^2}{\sigma^2}\cdot 0 = 0\,.$$

En consecuencia,

$$\lim_{n\to\infty}\varphi_{Z_n}(t) = e^{-\frac{t^2}{2}} = \varphi_Z(t)\,,$$

como queríamos demostrar. ♣

La rapidez de la convergencia que establece el teorema 13.8 depende de que la distribución de las variables $X_1, \ldots, X_n$ se *parezca*, o no, a la de una distribución normal. Obsérvese que, si son normales, el resultado es cierto sea cual sea el valor de n. Si las variables tienen una distribución acampanada y más o menos simétrica, valores de n del orden de 5 o 6 bastan para que la función de distribución de $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ sea prácticamente indistinguible de la normal estándar. Si la distribución es poco simétrica, pero sigue siendo razonablemente acampanada y unimodal, pueden conseguirse buenas aproximaciones para valores de n entre 20 y 30. Pero si la distribución de las variables es de las que se denominan en U (pequeñas probabilidades en el centro y mucho mayores en los extremos), pueden ser necesarios valores de n superiores a 50, e incluso a 100, para conseguir una aproximación razonable de las funciones de distribución.

13.3.1 Consecuencias del teorema del límite central

1) Muchas variables aleatorias son el resultado de la suma de un gran número de variables aleatorias *más o menos* independientes y *más o menos* igualmente distribuidas. El teorema del límite central permite afirmar que estas variables seguirán *más o menos* una distribución normal.

Así, el comportamiento aleatorio de pesos, alturas, consumos eléctricos, rentas, errores aleatorios, etc., se ajusta razonablemente a una distribución normal. En la altura de una persona (como en la mayoría de las variables biológicas) intervienen muchos factores: herencia genética, entorno, alimentación, constumbres de vida, etc. El consumo eléctrico diario es la suma de los consumos individuales en los cuales, a su vez, influyen muy diversos factores. La suposición de normalidad para este tipo de variables encuentra su apoyo teórico en el teorema del límite central.

En general, muchas de las variables económicas se forman por agregación de una infinidad de efectos individuales, por lo que la hipótesis de normalidad para dichas variables no es más que una consecuencia de este importante teorema.

2) El teorema del límite central afirma que, si las variables de la sucesión $\{X_n\}_{n=1}^{\infty}$ son independientes e igualmente distribuidas, de media μ y varianza σ^2, la sucesión $\{Y_n\}_{n=1}^{\infty}$, donde

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}},$$

converge en ley hacia una variable con distribución normal. Ahora bien, Y_n puede expresarse como

$$Y_n = \frac{n\left(\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right)}{n\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}},$$

siendo $E(\overline{X}_n) = \mu$ y $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Por tanto,

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{L}} N(0,1) ,$$

o lo que es lo mismo, $\overline{X}_n$ se distribuye asintóticamente como una $N(\mu, \frac{\sigma}{\sqrt{n}})$, es decir[5],

$$\overline{X}_n \overset{asint.}{\hookrightarrow} N(\mu, \frac{\sigma}{\sqrt{n}}) .$$

En consecuencia, cuando n sea grande, la media muestral, $\overline{X}_n$, se distribuirá aproximadamente como una $N(\mu, \frac{\sigma}{\sqrt{n}})$. Nótese que, si las variables X_i son $N(\mu, \sigma)$, la distribución exacta y asintótica coinciden, pues entonces,

$$\overline{X}_n \hookrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) .$$

Esto no ocurre en otro tipo de variables. Así, si las variables X_i siguen una distribución $\gamma(a,p)$, la distribución exacta de $\overline{X}_n$ es una $\gamma(na, np)$, de acuerdo con el corolario 10.8.1. Por otro lado, si aplicamos el teorema del límite central, obtendremos que $\overline{X}_n$ se distribuye aproximadamente como una $N(\frac{p}{a}, \frac{\sqrt{p}}{a\sqrt{n}})$. Entonces, para n grande, la distribución $\gamma(na, np)$ será parecida a la distribución $N(\frac{p}{a}, \frac{\sqrt{p}}{a\sqrt{n}})$. En este caso, se pueden obtener tanto la distribución exacta como la aproximada, aunque ambas no coinciden.

Finalmente, recordemos que, cuando X_i $(i = 1, \ldots, n)$ son variables independientes de Cauchy estándar, la media muestral $\overline{X}_n$ sigue, también, una distribución de Cauchy estándar que no depende de n (proposición 10.13). Por tanto, cuando n crece, la distribución de $\overline{X}_n$ no se irá aproximando a una normal. Este hecho no contradice el teorema del límite central, pues dicho teorema no puede aplicarse a las variables de Cauchy que carecen de momentos de primer y segundo orden.

3) El teorema del límite central permite aproximar distintas distribuciones a la normal; veámoslo en algunos caso particulares:

a) Si aplicamos el teorema a variables de Bernoulli (el denominado teorema de Moivre-Laplace), es decir, si la sucesión $\{X_n\}_{n=1}^{\infty}$ está formada por variables de Bernoulli, $b(p)$, independientes, entonces,

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{npq}} \xrightarrow{\mathcal{L}} N(0,1) ,$$

o lo que es lo mismo,

$$\sum_{i=1}^{n} X_i \overset{asint.}{\hookrightarrow} N(np, \sqrt{npq}).$$

Por tanto, teniendo en cuenta que la suma de variables de Bernoulli de parámetro p e independientes es una binomial,

$$\mathcal{B}(n,p) \overset{asint.}{\hookrightarrow} N(np, \sqrt{npq}) .$$

5. Adviértase que la siguiente expresión no significa que $\overline{X}_n$ converja en ley hacia una $N(\mu, \frac{\sigma}{\sqrt{n}})$, dado que esta última distribución depende de n.

b) Aplicando el teorema a variables de Poisson, $\mathcal{P}(\lambda)$, independientes, con $E(X_i) = Var(X_i) = \lambda$, entonces,

$$\frac{\sum_{i=1}^{n} X_i - n\lambda}{\sqrt{n\lambda}} \xrightarrow{\mathcal{L}} N(0,1) .$$

Pero, como $\sum_{i=1}^{n} X_i$ sigue una distribución de Poisson de parámetro $n \cdot \lambda$, resulta que

$$\mathcal{P}(n\lambda) \overset{asint.}{\hookrightarrow} N(n\lambda, \sqrt{n\lambda}) .$$

c) Si aplicamos el teorema a variables con distribución χ^2 con m grados de libertad e independientes (por tanto, $E(X_i) = m$ y $Var(X_i) = 2m$), entonces,

$$\frac{\sum_{i=1}^{n} X_i - nm}{\sqrt{2nm}} \xrightarrow{\mathcal{L}} N(0,1) .$$

Pero $\sum_{i=1}^{n} X_i$ se distribuye como una χ^2 con $n \cdot m$, por lo cual,

$$\chi^2_{nm} \overset{asint.}{\hookrightarrow} N(nm, \sqrt{2nm}) .$$

Igualmente encontramos resultados análogos para otros tipos de variables. Todos estos resultados asintóticos son aplicados por aproximación para ciertos valores de los parámetros; en términos generales, las aproximaciones son buenas cuando el parámetro respecto del cual la distribución es asintótica es mayor que 30.

13.3.2 Corrección por continuidad de Fisher

Cuando se aproxima la distribución de probabilidades de una variable discreta mediante una distribución continua, surge un problema de cierta relevancia práctica a la hora de calcular probabilidades. El problema consiste, básicamente, en que las probabilidades aproximadas no están determinadas unívocamente. Ilustrémoslo con un ejemplo, y demos la solución que se adopta, denominada corrección por continuidad.

Supongamos que se realizan 50 lanzamientos independientes de una moneda perfecta y deseamos calcular la probabilidad de *obtener al menos 20 caras, pero no más de 30.* La distribución de X, *número de caras obtenidas*, es binomial, $\mathcal{B}(50, \frac{1}{2})$, pero la probabilidad pedida,

$$p\,[20 \leq X \leq 30] ,$$

cuyo cálculo directo es tedioso, no se encuentra en las tablas, que sólo suministran probabilidades binomiales para $n \leq 20$.

Como estamos en las condiciones que permiten utilizar el teorema del límite central, podemos construir una sucesión de variables aleatorias $\{X_n\}_{n=1}^{\infty}$, donde X_n vale la unidad, si en el lanzamiento n-ésimo se obtiene una cara, y cero si en el mismo se obtiene una cruz. Entonces, el *número de caras en los primeros 50 lanzamientos* será $\sum_{n=1}^{50} X_n$.

El teorema del límite central (en concreto, su aplicación a variables de Bernoulli, es decir, el llamado teorema de Moivre-Laplace), indica que la variable aleatoria

$$\frac{\sum_{i=1}^{50} X_i - 50 \cdot \frac{1}{2}}{\sqrt{50 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = \frac{\sum_{i=1}^{50} X_i - 25}{\sqrt{12.5}}$$

tiene aproximadamente una distribución normal estándar y, por tanto,

$$\begin{aligned} p\left[20 \leq \sum_{i=1}^{50} X_i \leq 30\right] &= p\left[\frac{20-25}{\sqrt{12.5}} \leq \frac{\sum_{i=1}^{50} X_i - 25}{\sqrt{12.5}} \leq \frac{30-25}{\sqrt{12.5}}\right] \approx \\ &\approx p[-1.4142 \leq Z \leq 1.4142] = 0.8427\,, \end{aligned}$$

donde Z es una variable aleatoria con distribución normal estándar.

Ahora bien, el suceso

$$\left[20 \leq \sum_{i=1}^{50} X_i \leq 30\right]$$

coincide con el suceso

$$\left[19.2 \leq \sum_{i=1}^{50} X_i \leq 30.4\right]\,,$$

y, en consecuencia,

$$p\left[20 \leq \sum_{i=1}^{50} X_i \leq 30\right] = p\left[19.2 \leq \sum_{i=1}^{50} X_i \leq 30.4\right]\,.$$

Si aplicamos el teorema del límite central para calcular esta última probabilidad,

$$\begin{aligned} p\left[19.2 \leq \sum_{i=1}^{50} X_i \leq 30.4\right] &= p\left[\frac{19.2-25}{\sqrt{12.5}} \leq \frac{\sum_{i=1}^{50} X_i - 25}{\sqrt{12.5}} \leq \frac{30.4-25}{\sqrt{12.5}}\right] \approx \\ &\approx p[-1.6405 \leq Z \leq 1.5273] = 0.8862\,. \end{aligned}$$

Dispondríamos, entonces, de una doble aproximación de la probabilidad de que el número de caras se encuentre entre 20 y 30. Podríamos obtener otras aproximaciones, puesto que la probabilidad

$$p\left[20 \leq \sum_{i=1}^{50} X_i \leq 30\right]$$

coincide con la de cualquier suceso de la forma

$$\left[a \leq \sum_{i=1}^{50} X_i \leq b\right]\,,$$

siempre que $19 < a \leq 20$ y $30 \leq b < 31$.

El problema que ilustra este ejemplo se plantea siempre que aproximamos la distribución de una variable discreta mediante una distribución de probabilidades continua, y proviene del hecho de que dos valores consecutivos de una variable discreta están separados por un intervalo de valores que, si bien no son alcanzados por dicha variable, son valores posibles de la variable continua que aproxima su distribución.

La solución más correcta desde el punto de vista de los cálculos se debe a FISHER, y consiste en imaginar que la variable discreta resulta de agrupar en clases la variable continua, de forma que los valores de aquélla serían las marcas de clase (punto representativo del intervalo de valores que constituye una clase, en concreto, el punto medio de dicho intervalo). Si, por ejemplo, la variable discreta X toma únicamente valores enteros positivos, el suceso $X = n$ se debe escribir como

$$\left[n - \frac{1}{2} < X \leq n + \frac{1}{2}\right] ,$$

para cualquier n y, por tanto, a la hora de aproximar la probabilidad de un suceso de la forma $[n \leq X \leq m]$ mediante una distribución continua, dicho suceso deberá escribirse como

$$\left[n - \frac{1}{2} < X \leq m + \frac{1}{2}\right] .$$

Si aplicamos esta corrección por continuidad a nuestro ejemplo, la probabilidad buscada será

$$p\left[20 \leq \sum_{i=1}^{50} X_i \leq 30\right] = p\left[19.5 < \sum_{i=1}^{50} X_i \leq 30.5\right] ,$$

que, calculada mediante la aproximación normal, resultará:

$$\begin{aligned} p\left[19.5 < \sum_{i=1}^{50} X_i \leq 30.5\right] &= p\left[\frac{19.5 - 25}{\sqrt{12.5}} < \frac{\sum_{i=1}^{50} X_i - 25}{\sqrt{12.5}} \leq \frac{30.5 - 25}{\sqrt{12.5}}\right] \approx \\ &\approx p[-1.5556 \leq Z \leq 1.5556] = 0.8802 . \end{aligned}$$

Esta idea se generaliza, sin dificultad, al caso en que la variable discreta tome valores que no sean enteros. Si X es una variable discreta y la probabilidad del suceso $[X = x_i]$ va a calcularse por una aproximación continua, dicho suceso debe escribirse como

$$\left[\frac{x_{i-1} + x_i}{2} < X \leq \frac{x_i + x_{i+1}}{2}\right] .$$

En consecuencia, la probabilidad de un suceso de la forma

$$[x_i \leq X \leq x_j] ,$$

con $i < j$, se calculará escribiéndolo como

$$\left[\frac{x_{i-1} + x_i}{2} < X \leq \frac{x_j + x_{j+1}}{2}\right] .$$

Obsérvese que la desigualdad de la izquierda es estricta, mientras que la de la derecha no lo es. El objetivo de esta precisión es que los distintos intervalos no se solapen, aunque las probabilidades son las mismas, incluyan o no los intervalos a los extremos.

Ejemplo 13.8. El promedio de bolígrafos que se venden diariamente en una papelería es 30, siendo la desviación típica 5. Estos valores son 20 y 4, para el número de cuadernos vendidos. Se sabe, además, que el coeficiente de correlación entre las ventas de ambos productos es 0.7. ¿Cuál es la probabilidad de que el número total de ambos artículos vendidos durante un trimestre esté comprendido entre 4300 y 4600 unidades?

Sean X_{1i} y X_{2i} las variables aleatorias que designan, respectivamente, el número de bolígrafos y el número de cuadernos vendidos el i-ésimo día ($i = 1, \ldots, 90$). Sabemos que $E(X_{1i}) = 30$, $E(X_{2i}) = 20$, $Var(X_{1i}) = 25$, $Var(X_{2i}) = 16$ y que el coeficiente de correlación entre X_{1i} y X_{2i} es 0.7.

Si denotamos por Y_i la variable aleatoria *número total de unidades vendidas el día i-ésimo*, evidentemente,

$$Y_i = X_{1i} + X_{2i} \, .$$

Las variables aleatorias $Y_1, \ldots, Y_{90}$ son independientes e igualmente distribuidas, con esperanza

$$E(Y_i) = E(X_{1i}) + E(X_{2i}) = 30 + 20 = 50$$

y varianza

$$Var(Y_i) = Var(X_{1i}) + Var(X_{2i}) + 2 \cdot Cov(X_{1i}, X_{2i}) = 69 \, ,$$

ya que $Cov(X_{1i}, X_{2i})$, calculada a partir del coeficiente de correlación y de las desviaciones, vale 14.

La probabilidad pedida, esto es,

$$p\left[4300 < \sum_{i=1}^{90} Y_i \leq 4600\right] \, ,$$

se puede calcular por aproximación, sin más que aplicar el teorema del límite central; es decir, considerando que $\sum_{i=1}^{90} Y_i$, cuya distribución exacta es desconocida, sigue aproximadamente una ditribución normal $N(90 \cdot 50, \sqrt{69} \cdot \sqrt{90})$, o, lo que es lo mismo, una $N(4500, 78.8)$.

Ahora bien, antes de hallar dicha probabilidad hemos de tener en cuenta que el hecho de aproximar una distribución discreta a una continua obliga a utilizar la corrección por continuidad de Fisher. Consecuentemente, la probabilidad que habrá que buscar será

$$p\left[4300.5 < \sum_{i=1}^{90} Y_i \leq 4600.5\right]$$

que, tipificando, resulta

$$p\left[4300.5 < \sum_{i=1}^{90} Y_i \leq 4600.5\right] = p\left[\frac{4300.5 - 4500}{78.8} < \frac{\sum_{i=1}^{90} Y_i - 4500}{78.8} \leq \frac{4600.5 - 4500}{78.8}\right] .$$

Aproximando por la distribución normal,

$$p\left[4300.5 < \sum_{i=1}^{90} Y_i \leq 4600.5\right] = p[-2.53 < Z \leq 1.28] = 0.894 \, ,$$

donde $Z \hookrightarrow N(0,1)$. Por tanto, la probabilidad de vender trimestralmente entre 4300 y 4600 bolígrafos y cuadernos es, aproximadamente, 0.894. $\triangleleft$

13.4 Ejercicios propuestos

13.1. Demuéstrese que si una sucesión $\{X_n\}_{n=1}^{\infty}$ converge casi seguro hacia X, la sucesión $\{X_n - X\}_{n=1}^{\infty}$ converge casi seguro hacia cero.

13.2. Demuéstrese que si una sucesión $\{X_n\}_{n=1}^{\infty}$ converge en probabilidad hacia X, la sucesión $\{X_n - X\}_{n=1}^{\infty}$ converge en probabilidad hacia cero.

13.3. Sean las sucesiones de variables aleatorias, $\{X_n\}_{n=1}^{\infty}$, definidas como:

a)

$$X_n : \left| \begin{array}{ll} 0 & p[X_n = 0] = 1 - \frac{1}{n} \\ n & p[X_n = n] = \frac{1}{n} \end{array} \right.$$

b)

$$X_n : \left| \begin{array}{ll} 0 & p[X_n = 0] = 1 - \frac{1}{n} \\ 1 & p[X_n = 1] = \frac{1}{n} \end{array} \right.$$

Estúdiense las convergencias de las sucesiones hacia el cero. (Sugerencia: la serie $\sum_{n=1}^{\infty} \frac{1}{n}$ es divergente.)

13.4. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias tales que

$$X_n : \left| \begin{array}{ll} 1 & p[X_n = 1] = \frac{1}{n^2} \\ 0 & p[X_n = 0] = 1 - \frac{1}{n^2} \end{array} \right.$$

Estúdiese la convergencia en probabilidad de la sucesión hacia el cero.

13.5. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias tales que

$$X_n : \left| \begin{array}{ll} \frac{1}{n} & p\left[X_n = \frac{1}{n}\right] = \frac{n^2-1}{n^2} \\ n & p[X_n = n] = \frac{1}{n^2} \end{array} \right.$$

Estúdiese la convergencia casi segura, en media y en ley de la sucesión hacia el cero.

13.6. Se consideran lanzamientos consecutivos e independientes de una moneda. Sea la variable aleatoria:

$$X_n : \left| \begin{array}{ll} \frac{1}{n} & \text{si el } n\text{-ésimo resultado es cara} \\ 0 & \text{si el } n\text{-ésimo resultado es cruz} \end{array} \right.$$

Estúdiense las distintas convergencias de la sucesión $\{X_n\}_{n=1}^{\infty}$.

13.7. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias tales que X_n sigue una distribución $\mathcal{B}(n, \frac{\lambda}{n})$, y sea X una variable aleatoria con distribución de Poisson, $\mathcal{P}(\lambda)$. Sean φ_n y φ las correspondientes funciones características. Demuéstrese que

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t) \, .$$

13.8. Sea $\{(X_n, Y_n)\}_{n=1}^{\infty}$ una sucesión de variables aleatorias bidimensionales, independientes dos a dos e igualmente distribuidas, existiendo $\alpha_{2,0}$, $\alpha_{0,2}$ y $\alpha_{1,1}$. Demuéstrese que

$$S_{XY} \xrightarrow{c.p.} \sigma_{XY}$$

y

$$R_{XY} \xrightarrow{c.p.} \rho_{XY} \, ,$$

donde S_{XY} y R_{XY} son la covarianza y el coeficiente de correlación muestrales y σ_{XY} y ρ_{XY} las correspondientes características poblacionales.

13.9. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias independientes e igualmente distribuidas tales que

$$X_n : \left| \begin{array}{ll} 1 & p[X_n = 1] = \frac{1}{2} \\ -1 & p[X_n = -1] = \frac{1}{2} \end{array} \right.$$

a) Demuéstrese que $\varphi_{X_n}(t) = \cos(t)$.

b) Compruébese que

$$\lim_{n\to\infty} \varphi_{X_n}(t) = \varphi_Z(t) \, ,$$

siendo $\varphi_Z(t)$ la función característica de una normal estándar, es decir, que $\{X_n\}_{n=1}^{\infty}$ verifica el teorema del límite central. (Sugerencia: recuérdese que $1 - \cos(x) = 2\text{sen}^2(\frac{x}{2})$.)

13.10. El número de erratas de imprenta en una página de un libro sigue una distribución de Poisson, siendo 2 el número medio de erratas por página. En un libro de 300 páginas, ¿cuál es la probabilidad de que en una o más páginas existan más de 5 erratas?

13.11. Una máquina empaqueta lotes de folios. El dueño de la máquina desea que al menos el 90 % de dichos lotes tengan más de mil folios no defectuosos. Sabiendo que la probabilidad de que un folio sea defectuoso es 0.02, ¿qué número de folios deberá disponer por lote?

13.12. En un museo se venden diariamente mil entradas, siendo la proporción de visitantes extranjeros del 35 %. ¿Cuál es la probabilidad de que en una semana fueran más de 5 mil los españoles que visitaron este museo?

13.13. Las galletas Gruñón están de oferta: algunos de sus paquetes incluyen vales canjeables por viajes al Caribe.
El número de paquetes premiados que se venden al día en un establecimiento es una variable de Poisson de parámetro 0.5. Calcúlese la probabilidad de que en 100 días se vendan más de 10 paquetes con premio.

13.14. La renta de los individuos de un país sigue una distribución de Pareto con $x_0 = 70000$ pesetas y $\alpha = 0.5$. ¿Cuál es la probabilidad de que la renta media de un grupo de 1000 individuos sea superior a 100000 pesetas?

13.15. Tres cadenas de televisión tienen una dura competencia por conseguir los más altos niveles de audiencia.
El promedio diario de premios millonarios es de 5, 3 y 4 para cada una de ellas, siendo las desviaciones típicas de 0.5, 0.4 y 0.3, respectivamente.
Determínese la probabilidad de que el número total de premios millonarios repartidos en dos meses sea superior a 400.

13.16. En las últimas semanas, el servicio de urgencias de una hospital se encuentra desbordado de trabajo, pues el número de pacientes que requieren este tipo de atención ha aumentado de modo alarmante.
Si el número medio de gasas que se utilizan diariamente en este servicio es de 200, con una desviación típica de 20, hállese el número mínimo de gasas que deberá encargar el supervisor de esta unidad para asegurar con probabilidad del 99 % que todas las necesidades queden cubiertas durante un período de medio año.

13.17. Cincuenta estudiantes de quinto curso de una facultad decidieron vender camisetas para costear su viaje de fin de carrera. Como no estaban muy seguros del éxito que iban a tener, adquirieron pocas camisetas; sin embargo, vendieron todas las unidades en tan poco tiempo que optaron por comprar más. Sabiendo que el promedio de camisetas vendidas por alumno es de 30 y la desviación típica de 6:

a) ¿Cuál es el número máximo de camisetas que deben encargar para asegurar la venta total con una probabilidad del 90 %?

b) Si por cada camiseta se obtiene un beneficio de 500 pesetas, ¿cuál es la probabilidad de que el beneficio total para esta última remesa de camisetas supere el medio millón de pesetas?

ANEXO
Tablas estadísticas

I. Función de distribución binomial

La función tabulada para $n \leq 20$ y $p \leq 0.5$ es:

$$F_X(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \qquad \text{para } k = 0, 1, \ldots, n\,.$$

Si $p > 0.5$, utilizaremos el resultado $F_X(k) = 1 - F_Y(n-k-1)$, siendo X e Y variables con distribución binomial de parámetros (n,p) y $(n, 1-p)$, respectivamente.

La función de probabilidad es:

$$\begin{aligned} p[X=0] &= F_X(0) \\ p[X=k] &= F_X(k) - F_X(k-1) \qquad \text{para } k = 1, \ldots, n. \end{aligned}$$

Si $n > 20$, la distribución binomial se aproxima a una distribución normal de media np y varianza npq.

Si $n > 20$, siendo p pequeño y $np \leq 20$, la distribución binomial se aproxima a una distribución de Poisson de media np.

		p											
n	k	0,01	0,05	0,10	0,15	0,20	0,25	0,30	1/3	0,35	0,40	0,45	0,50
2	0	0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4444	0,4225	0,3600	0,3025	0,2500
	1	0,9999	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8889	0,8775	0,8400	0,7975	0,7500
	2	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
3	0	0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2963	0,2746	0,2160	0,1664	0,1250
	1	0,9997	0,9928	0,9720	0,9392	0,8960	0,8438	0,7840	0,7407	0,7182	0,6480	0,5748	0,5000
	2	1,0000	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9630	0,9571	0,9360	0,9089	0,8750
			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	0	0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1975	0,1785	0,1296	0,0915	0,0625
	1	0,9994	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5926	0,5630	0,4752	0,3910	0,3125
	2	1,0000	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8889	0,8735	0,8208	0,7585	0,6875
	3		1,0000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9876	0,9850	0,9744	0,9590	0,9375
	4			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
5	0	0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1317	0,1160	0,0778	0,0503	0,0312
	1	0,9990	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4609	0,4284	0,3370	0,2562	0,1875
	2	1,0000	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7901	0,7648	0,6826	0,5931	0,5000
	3		1,0000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9547	0,9460	0,9130	0,8688	0,8125
	4			1,0000	0,9999	0,9997	0,9990	0,9976	0,9959	0,9947	0,9898	0,9815	0,9688
	5				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
6	0	0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0878	0,0754	0,0467	0,0277	0,0156
	1	0,9986	0,9672	0,8857	0,7765	0,6553	0,5339	0,4202	0,3512	0,3191	0,2333	0,1636	0,1094
	2	1,0000	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6804	0,6471	0,5443	0,4415	0,3438
	3		0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,8999	0,8826	0,8208	0,7447	0,6562
	4		1,0000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9822	0,9777	0,9590	0,9308	0,8906
	5			1,0000	1,0000	0,9999	0,9998	0,9993	0,9986	0,9982	0,9959	0,9917	0,9844
	6					1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

FUNCIÓN DE LA DISTRIBUCIÓN BINOMIAL *(cont.)*

		p											
n	*k*	0,01	0,05	0,10	0,15	0,20	0,25	0,30	1/3	0,35	0,40	0,45	0,50
7	0	0,9321	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0585	0,0490	0,0280	0,0152	0,0078
	1	0,9980	0,9556	0,8503	0,7166	0,5767	0,4449	0,3294	0,2634	0,2338	0,1586	0,1024	0,0625
	2	1,0000	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5706	0,5323	0,4199	0,3164	0,2266
	3		0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8267	0,8002	0,7102	0,6083	0,5000
	4		1,0000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9547	0,9444	0,9037	0,8471	0,7734
	5			1,0000	0,9999	0,9996	0,9987	0,9962	0,9931	0,9910	0,9812	0,9643	0,9375
	6				1,0000	1,0000	0,9999	0,9998	0,9995	0,9994	0,9984	0,9963	0,9922
	7						1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
8	0	0,9227	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0390	0,0319	0,0168	0,0084	0,0039
	1	0,9973	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1951	0,1691	0,1064	0,0632	0,0352
	2	0,9999	0,9942	0,9619	0,8948	0,7969	0,6785	0,5518	0,4682	0,4278	0,3154	0,2201	0,1445
	3	1,0000	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7413	0,7064	0,5941	0,4770	0,3633
	4		1,0000	0,9996	0,9971	0,9896	0,9727	0,9420	0,9121	0,8939	0,8263	0,7396	0,6367
	5			1,0000	0,9998	0,9988	0,9958	0,9887	0,9803	0,9747	0,9502	0,9115	0,8555
	6				1,0000	0,9999	0,9996	0,9987	0,9974	0,9964	0,9915	0,9819	0,9648
	7					1,0000	1,0000	0,9999	0,9998	0,9998	0,9993	0,9983	0,9961
	8							1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
9	0	0,9135	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0260	0,0207	0,0101	0,0046	0,0020
	1	0,9965	0,9288	0,7748	0,5995	0,4362	0,3003	0,1960	0,1431	0,1211	0,0705	0,0385	0,0195
	2	0,9999	0,9916	0,9470	0,8591	0,7382	0,6007	0,4628	0,3772	0,3373	0,2318	0,1495	0,0898
	3	1,0000	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6503	0,6089	0,4826	0,3614	0,2539
	4		1,0000	0,9991	0,9944	0,9804	0,9511	0,9012	0,1552	0,8283	0,7334	0,6214	0,5000
	5			0,9999	0,9994	0,9969	0,9900	0,9747	0,9576	0,9464	0,9006	0,8342	0,7461
	6			1,0000	1,0000	0,9997	0,9987	0,9957	0,9917	0,9888	0,9750	0,9502	0,9102
	7					1,0000	0,9999	0,9996	0,9990	0,9986	0,9962	0,9909	0,9805
	8						1,0000	1,0000	0,9999	0,9999	0,9997	0,9992	0,9980
	9								1,0000	1,0000	1,0000	1,0000	1,0000
10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0173	0,0135	0,0060	0,0025	0,0010
	1	0,9958	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,1040	0,0860	0,0464	0,0233	0,0107
	2	1,0000	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2991	0,2616	0,1673	0,0996	0,0547
	3		0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5593	0,5138	0,3823	0,2660	0,1719
	4		0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7869	0,7515	0,6331	0,5044	0,3770
	5		1,0000	0,9999	0,9986	0,9936	0,9803	0,9527	0,9234	0,9051	0,8338	0,7384	0,6230
	6			1,0000	0,9999	0,9991	0,9965	0,9894	0,8803	0,9740	0,9452	0,8980	0,8281
	7				1,0000	0,9999	0,9996	0,9984	0,9966	0,9952	0,9877	0,9726	0,9453
	8					1,0000	1,0000	0,9999	0,9996	0,9995	0,9983	0,9955	0,9893
	9							1,0000	1,0000	1,0000	0,9999	0,9997	0,9990
	10										1,0000	1,0000	1,0000
11	0	0,8954	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0116	0,0088	0,0036	0,0014	0,0005
	1	0,9948	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0752	0,0606	0,0302	0,0139	0,0059
	2	0,9998	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2341	0,2001	0,1189	0,0652	0,0327
	3	1,0000	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4726	0,4256	0,2963	0,1911	0,1133
	4		0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,7110	0,6683	0,5328	0,3971	0,2744
	5		1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8779	0,8513	0,7535	0,6331	0,5000
	6			1,0000	0,9997	0,9980	0,9924	0,9784	0,9614	0,9499	0,9006	0,8262	0,7256
	7				1,0000	0,9998	0,9988	0,9957	0,9912	0,9878	0,9707	0,9390	0,8867
	8					1,0000	0,9999	0,9994	0,9986	0,9980	0,9941	0,9852	0,9673
	9						1,0000	1,0000	0,9999	0,9998	0,9993	0,9978	0,9941
	10								1,0000	1,0000	1,0000	0,9998	0,9995
	11											1,0000	1,0000

FUNCIÓN DE LA DISTRIBUCIÓN BINOMIAL *(cont.)*

		p											
n	*k*	0,01	0,05	0,10	0,15	0,20	0,25	0,30	1/3	0,35	0,40	0,45	0,50
12	0	0,8864	0,5404	0,2824	0,1422	0,0687	0,0317	0,0138	0,0077	0,0057	0,0022	0,0008	0,0002
	1	0,9938	0,8816	0,6590	0,4435	0,2749	0,1584	0,0850	0,0540	0,0424	0,0196	0,0083	0,0032
	2	0,9998	0,9804	0,8891	0,7358	0,5583	0,3907	0,2528	0,1811	0,1513	0,0834	0,0421	0,0193
	3	1,0000	0,9978	0,9744	0,9078	0,7946	0,6488	0,4925	0,3931	0,3467	0,2253	0,1345	0,0730
	4		0,9998	0,9957	0,9761	0,9274	0,8424	0,7237	0,6315	0,5833	0,4382	0,3044	0,1938
	5		1,0000	0,9995	0,9954	0,9806	0,9456	0,8822	0,8223	0,7873	0,6652	0,5269	0,3872
	6			0,9999	0,9993	0,9961	0,9857	0,9614	0,9336	0,9154	0,8418	0,7393	0,6128
	7			1,0000	0,9999	0,9994	0,9972	0,9905	0,9812	0,9745	0,9427	0,8883	0,8062
	8				1,0000	0,9999	0,9996	0,9983	0,9961	0,9944	0,9847	0,9644	0,9270
	9					1,0000	1,0000	0,9998	0,9995	0,9992	0,9972	0,9921	0,9807
	10							1,0000	0,9999	0,9999	0,9997	0,9989	0,9968
	11								1,0000	1,0000	1,0000	0,9999	0,9998
	12											1,0000	1,0000
13	0	0,8775	0,5133	0,2542	0,1209	0,0550	0,0238	0,0097	0,0051	0,0037	0,0013	0,0004	0,0001
	1	0,9928	0,8646	0,6213	0,3983	0,2336	0,1267	0,0637	0,0358	0,0296	0,0126	0,0049	0,0017
	2	0,9997	0,9755	0,8661	0,6920	0,5017	0,3326	0,2025	0,1387	0,1132	0,0579	0,0269	0,0112
	3	1,0000	0,9969	0,9658	0,8820	0,7473	0,5843	0,4206	0,3224	0,2783	0,1686	0,0929	0,0461
	4		0,9997	0,9935	0,9658	0,9009	0,7940	0,6543	0,5520	0,5005	0,3530	0,2279	0,1334
	5		1,0000	0,9991	0,9924	0,9700	0,9198	0,8346	0,7587	0,7159	0,5744	0,4268	0,2905
	6			0,9999	0,9987	0,9930	0,9757	0,9376	0,8965	0,8705	0,7712	0,6437	0,5000
	7			1,0000	0,9998	0,9988	0,9944	0,9818	0,9654	0,9538	0,9023	0,8212	0,7095
	8				1,0000	0,9998	0,9990	0,9960	0,9912	0,9874	0,9679	0,9302	0,8666
	9					1,0000	0,9999	0,9993	0,9984	0,9975	0,9922	0,9797	0,9539
	10						1,0000	0,9999	0,9998	0,9997	0,9987	0,9959	0,9888
	11							1,0000	1,0000	1,0000	0,9999	0,9995	0,9983
	12										1,0000	1,0000	0,9999
	13												1,0000
14	0	0,8687	0,4877	0,2288	0,1028	0,0440	0,0178	0,0068	0,0034	0,0024	0,0008	0,0002	0,0001
	1	0,9916	0,8470	0,5846	0,3567	0,1979	0,1010	0,0475	0,0274	0,0205	0,0081	0,0029	0,0009
	2	0,9997	0,9699	0,8416	0,6479	0,4481	0,2811	0,1608	0,1053	0,0839	0,0398	0,0170	0,0065
	3	1,0000	0,9958	0,9559	0,8535	0,6982	0,5213	0,3552	0,2612	0,2205	0,1243	0,0632	0,0287
	4		0,9996	0,9908	0,9533	0,8702	0,7415	0,5842	0,4755	0,4227	0,2793	0,1672	0,0898
	5		1,0000	0,9985	0,9885	0,9561	0,8883	0,7805	0,6899	0,6405	0,4859	0,3373	0,2120
	6			0,9998	0,9978	0,9884	0,9617	0,9067	0,8505	0,8164	0,6925	0,5461	0,3953
	7			1,0000	0,9997	0,9976	0,9897	0,9685	0,9424	0,9247	0,8499	0,7414	0,6047
	8				1,0000	0,9996	0,9978	0,9917	0,9826	0,9757	0,9417	0,8811	0,7880
	9					1,0000	0,9997	0,9983	0,9960	0,9940	0,9825	0,9574	0,9102
	10						1,0000	0,9998	0,9993	0,9989	0,9961	0,9886	0,9713
	11							1,0000	0,9999	0,9999	0,9994	0,9978	0,9935
	12								1,0000	1,0000	0,9999	0,9997	0,9991
	13										1,0000	1,0000	0,9999
	14												1,0000

FUNCIÓN DE LA DISTRIBUCIÓN BINOMIAL *(cont.)*

		p											
n	*k*	0,01	0,05	0,10	0,15	0,20	0,25	0,30	1/3	0,35	0,40	0,45	0,50
15	0	0,8601	0,4633	0,2059	0,0874	0,0352	0,0134	0,0047	0,0023	0,0016	0,0005	0,0001	0,0000
	1	0,9904	0,8290	0,5490	0,3186	0,1671	0,0802	0,0353	0,0194	0,0142	0,0052	0,0017	0,0005
	2	0,9996	0,9638	0,8159	0,6042	0,3980	0,2361	0,1268	0,0794	0,0617	0,0271	0,0107	0,0037
	3	1,0000	0,9945	0,9444	0,8227	0,6482	0,4613	0,2969	0,2092	0,1727	0,0905	0,0424	0,0176
	4		0,9994	0,9873	0,9383	0,8358	0,6865	0,5155	0,4041	0,3519	0,2173	0,1204	0,0592
	5		0,9999	0,9978	0,9832	0,9389	0,8516	0,7216	0,3184	0,5643	0,4032	0,2608	0,1509
	6		1,0000	0,9997	0,9964	0,9819	0,9434	0,8689	0,8070	0,7548	0,6098	0,4522	0,3036
	7			1,0000	0,9994	0,9958	0,9827	0,9500	0,9118	0,8868	0,7869	0,6535	0,5000
	8				0,9999	0,9992	0,9958	0,9848	0,9692	0,9578	0,9050	0,8182	0,6964
	9				1,0000	0,9999	0,9992	0,9963	0,9915	0,9876	0,9662	0,9231	0,8491
	10					1,0000	0,9999	0,9993	0,9982	0,9972	0,9907	0,9745	0,9408
	11						1,0000	0,9999	0,9997	0,9995	0,9981	0,9937	0,9824
	12							1,0000	1,0000	0,9999	0,9997	0,9989	0,9963
	13									1,0000	1,0000	0,9999	0,9995
	14											1,0000	1,0000
16	0	0,8515	0,4401	0,1853	0,0743	0,0281	0,0100	0,0033	0,0015	0,0010	0,0003	0,0001	0,0000
	1	0,9891	0,8108	0,5147	0,2839	0,1407	0,0635	0,0261	0,0137	0,0098	0,0033	0,0010	0,0003
	2	0,9995	0,9571	0,7892	0,5614	0,3518	0,1971	0,0994	0,0594	0,0451	0,0183	0,0066	0,0021
	3	1,0000	0,9930	0,9316	0,7899	0,5981	0,4050	0,2459	0,1660	0,1339	0,0651	0,0281	0,0106
	4		0,9991	0,9830	0,9209	0,7982	0,6302	0,4499	0,3391	0,2892	0,1666	0,0853	0,0384
	5		0,9999	0,9967	0,9765	0,9183	0,8103	0,6598	0,5469	0,4900	0,3288	0,1976	0,1051
	6		1,0000	0,9995	0,9944	0,9733	0,9204	0,8247	0,7374	0,6881	0,5272	0,3660	0,2272
	7			0,9999	0,9989	0,9930	0,9729	0,9256	0,8735	0,8406	0,7161	0,5629	0,4018
	8			1,0000	0,9998	0,9985	0,9925	0,9743	0,9500	0,9329	0,8577	0,7441	0,5982
	9				1,0000	0,9998	0,9984	0,9929	0,9841	0,9771	0,9417	0,8759	0,7228
	10					1,0000	0,9997	0,9984	0,9960	0,9938	0,9809	0,9514	0,8949
	11						1,0000	0,9997	0,9992	0,9987	0,9951	0,9851	0,9616
	12							1,0000	0,9999	0,9998	0,9991	0,9965	0,9894
	13								1,0000	1,0000	0,9999	0,9994	0,9979
	14										1,0000	0,9999	0,9997
	15											1,0000	1,0000
17	0	0,8429	0,4181	0,1668	0,0631	0,0225	0,0075	0,0023	0,0010	0,0007	0,0002	0,0000	0,0000
	1	0,9877	0,7922	0,4818	0,2525	0,1182	0,0501	0,0193	0,0096	0,0067	0,0021	0,0006	0,0001
	2	0,9994	0,9497	0,7618	0,5198	0,3096	0,1637	0,0774	0,0442	0,0327	0,0123	0,0041	0,0012
	3	1,0000	0,9912	0,9174	0,7556	0,5489	0,3530	0,2019	0,1304	0,1028	0,0464	0,0184	0,0064
	4		0,9988	0,9779	0,9013	0,7582	0,5739	0,3887	0,2814	0,2348	0,1260	0,0596	0,0245
	5		0,9999	0,9953	0,9681	0,8943	0,7653	0,5968	0,6739	0,4197	0,2639	0,1471	0,0717
	6		1,0000	0,9992	0,9917	0,9623	0,8929	0,7752	0,8281	0,6188	0,4478	0,2902	0,1662
	7			0,9999	0,9983	0,9891	0,9598	0,8954	0,9245	0,7872	0,6405	0,4743	0,3145
	8			1,0000	0,9997	0,9974	0,9876	0,9597	0,9727	0,9006	0,8011	0,6626	0,5000
	9				1,0000	0,9995	0,9969	0,9873	0,9920	0,9617	0,9081	0,8166	0,6855
	10					0,9999	0,9994	0,9968	0,9981	0,9880	0,9652	0,9174	0,8338
	11					1,0000	0,9999	0,9993	0,9997	0,9970	0,9894	0,9699	0,9283
	12						1,0000	0,9999	0,9999	0,9994	0,9975	0,9914	0,9755
	13							1,0000	1,0000	0,9999	0,9995	0,9981	0,9936
	14									1,0000	0,9999	0,9997	0,9988
	15										1,0000	1,0000	0,9999
	16												1,0000

FUNCIÓN DE LA DISTRIBUCIÓN BINOMIAL *(cont.)*

		p											
n	*k*	0,01	0,05	0,10	0,15	0,20	0,25	0,30	1/3	0,35	0,40	0,45	0,50
18	0	0,8345	0,3972	0,1501	0,0536	0,0180	0,0056	0,0016	0,0007	0,0004	0,0001	0,0000	0,0000
	1	0,9862	0,7735	0,4503	0,2241	0,0991	0,0395	0,0142	0,0068	0,0046	0,0013	0,0003	0,0001
	2	0,9993	0,9419	0,7338	0,4797	0,2713	0,1353	0,0600	0,0326	0,0236	0,0082	0,0025	0,0007
	3	1,0000	0,9891	0,9018	0,7202	0,5010	0,3057	0,1646	0,1017	0,0783	0,0328	0,0120	0,0038
	4		0,9985	0,9718	0,8794	0,7164	0,5187	0,3327	0,2311	0,1886	0,0942	0,0411	0,0154
	5		0,9998	0,9936	0,9581	0,8671	0,7175	0,5344	0,4122	0,3550	0,2088	0,1077	0,0481
	6		1,0000	0,9988	0,9882	0,9487	0,8610	0,7217	0,6085	0,5491	0,3743	0,2258	0,1189
	7			0,9998	0,9973	0,9837	0,9431	0,8593	0,7767	0,7283	0,5634	0,3915	0,2403
	8			1,0000	0,9995	0,9957	0,9807	0,9404	0,8924	0,8609	0,7368	0,5778	0,4073
	9				0,9999	0,9991	0,9946	0,9790	0,9566	0,9403	0,8653	0,7473	0,5927
	10				1,0000	0,9998	0,9988	0,9939	0,9856	0,9788	0,9424	0,8720	0,7597
	11					1,0000	0,9998	0,9986	0,9961	0,9938	0,9797	0,9463	0,8811
	12						1,0000	0,9997	0,9992	0,9986	0,9942	0,9817	0,9519
	13							1,0000	0,9999	0,9997	0,9987	0,9951	0,9846
	14								1,0000	1,0000	0,9998	0,9990	0,9962
	15										1,0000	0,9999	0,9993
	16											1,0000	0,9999
	17												1,0000
19	0	0,8262	0,3774	0,1351	0,0456	0,0144	0,0042	0,0011	0,0004	0,0003	0,0001	0,0000	0,0000
	1	0,9847	0,7547	0,4203	0,1985	0,0829	0,0310	0,0104	0,0047	0,0031	0,0008	0,0002	0,0000
	2	0,9991	0,9335	0,7054	0,4413	0,2369	0,1113	0,0462	0,0240	0,0170	0,0055	0,0015	0,0004
	3	1,0000	0,9868	0,8850	0,6841	0,4551	0,2631	0,1332	0,0787	0,0591	0,0230	0,0077	0,0022
	4		0,9980	0,9648	0,8556	0,6733	0,4654	0,2822	0,1879	0,1500	0,0696	0,0280	0,0096
	5		0,9998	0,9914	0,9463	0,8369	0,6678	0,4739	0,3518	0,2968	0,1629	0,0777	0,0318
	6		1,0000	0,9983	0,9837	0,9324	0,8251	0,6655	0,5431	0,4812	0,3081	0,1727	0,0835
	7			0,9997	0,9959	0,9767	0,9225	0,8180	0,7207	0,6656	0,4878	0,3169	0,1796
	8			1,0000	0,9992	0,9933	0,9713	0,9161	0,8538	0,8145	0,6675	0,4940	0,3238
	9				0,9999	0,9984	0,9911	0,9674	0,9352	0,9125	0,8139	0,6710	0,5000
	10				1,0000	0,9997	0,9977	0,9895	0,9759	0,9653	0,9115	0,8159	0,6762
	11					1,0000	0,9995	0,9972	0,9926	0,9886	0,9648	0,9129	0,8204
	12						0,9999	0,9994	0,9981	0,9969	0,9884	0,9658	0,9165
	13						1,0000	0,9999	0,9996	0,9993	0,9969	0,9891	0,9682
	14							1,0000	0,9999	0,9999	0,9994	0,9972	0,9904
	15								1,0000	1,0000	0,9999	0,9995	0,9978
	16										1,0000	0,9999	0,9996
	17											1,0000	1,0000
20	0	0,8179	0,3585	0,1216	0,0388	0,0115	0,0032	0,0008	0,0003	0,0002	0,0000	0,0000	0,0000
	1	0,9831	0,7358	0,3917	0,1756	0,0692	0,0243	0,0076	0,0033	0,0021	0,0005	0,0001	0,0000
	2	0,9990	0,9245	0,6769	0,4049	0,2061	0,0913	0,0355	0,0176	0,0121	0,0036	0,0009	0,0002
	3	1,0000	0,9841	0,8670	0,6477	0,4114	0,2252	0,1071	0,0604	0,0444	0,0160	0,0049	0,0013
	4		0,9974	0,9568	0,8298	0,6296	0,4148	0,2375	0,1515	0,1182	0,0510	0,0189	0,0059
	5		0,9997	0,9887	0,9327	0,8042	0,6172	0,4164	0,2972	0,2454	0,1256	0,0553	0,0207
	6		1,0000	0,9976	0,9781	0,9133	0,7858	0,6080	0,4793	0,4166	0,2500	0,1299	0,0577
	7			0,9996	0,9941	0,9679	0,8982	0,7723	0,6615	0,6010	0,4159	0,2520	0,1316
	8			0,9999	0,9987	0,9900	0,9591	0,8867	0,8094	0,7624	0,5956	0,4143	0,2517
	9			1,0000	0,9998	0,9974	0,9861	0,9520	0,9081	0,8782	0,7553	0,5914	0,4119
	10				1,0000	0,9994	0,9961	0,9829	0,9624	0,9468	0,8725	0,7507	0,5881
	11					0,9999	0,9991	0,9949	0,9870	0,9804	0,9435	0,8692	0,7483
	12					1,0000	0,9998	0,9987	0,9963	0,9940	0,9790	0,9420	0,8684
	13						1,0000	0,9997	0,9991	0,9985	0,9935	0,9786	0,9423
	14							1,0000	0,9999	0,9997	0,9984	0,9936	0,9793
	15								1,0000	1,0000	0,9997	0,9985	0,9941
	16										1,0000	0,9997	0,9987
	17											1,0000	0,9998
	18												1,0000

II. Función de distribución de Poisson

La función tabulada para $\lambda \leq 20$ es:

$$F_X(k) = \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^i}{i!} \qquad \text{para } k = 0, 1, \ldots$$

La función de probabilidad es:

$$\begin{aligned} p[X = 0] &= F_X(0) \\ p[X = k] &= F_X(k) - F_X(k-1) \qquad \text{para } k = 1, 2, \ldots \end{aligned}$$

Si $\lambda > 20$, la distribución de Poisson se aproxima a una distribución normal de media y varianza λ.

	λ									
k	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810
4		1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963
5				1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994
6							1,0000	1,0000	1,0000	0,9999
7										1,0000

	λ									
k	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4338	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571
4	0,9946	0,9923	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473
5	0,9990	0,9985	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834
6	0,9999	0,9997	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955
7	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989
8			1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9998
9							1,0000	1,0000	1,0000	1,0000

	λ									
k	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3,0
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487	0,2311	0,2146	0,1991
2	0,6496	0,6227	0,5960	0,5697	0,5438	0,5184	0,4936	0,4695	0,4460	0,4232
3	0,8386	0,8194	0,7993	0,7787	0,7576	0,7360	0,7141	0,6919	0,6696	0,6472
4	0,9379	0,9275	0,9163	0,9041	0,8912	0,8774	0,8629	0,8477	0,8318	0,8153
5	0,9796	0,9751	0,9700	0,9643	0,9580	0,9510	0,9433	0,9349	0,9258	0,9161
6	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794	0,9756	0,9713	0,9665
7	0,9985	0,9980	0,9974	0,9967	0,9958	0,9947	0,9934	0,9919	0,9901	0,9881
8	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981	0,9976	0,9969	0,9962
9	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995	0,9993	0,9991	0,9989
10	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
11					1,0000	1,0000	1,0000	1,0000	0,9999	0,9999

FUNCIÓN DE LA DISTRIBUCIÓN POISSON *(cont.)*

	λ									
k	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0
0	0,0450	0,0408	0,0369	0,0334	0,0302	0,0273	0,0247	0,0224	0,0202	0,0183
1	0,1847	0,1712	0,1586	0,1468	0,1359	0,1257	0,1162	0,1074	0,0992	0,0916
2	0,4012	0,3799	0,3594	0,3397	0,3208	0,3027	0,2854	0,2689	0,2531	0,2381
3	0,6248	0,6025	0,5803	0,5584	0,5366	0,5152	0,4942	0,4735	0,4533	0,4335
4	0,7982	0,7806	0,7626	0,7442	0,7254	0,7064	0,6872	0,6678	0,6484	0,6288
5	0,9057	0,8946	0,8829	0,8705	0,8576	0,8441	0,8301	0,8156	0,8006	0,7851
6	0,9612	0,9554	0,9490	0,9421	0,9347	0,9267	0,9182	0,9091	0,8995	0,8893
7	0,9858	0,9832	0,9802	0,9769	0,9733	0,9692	0,9648	0,9599	0,9546	0,9489
8	0,9953	0,9943	0,9931	0,9917	0,9901	0,9883	0,9863	0,9840	0,9815	0,9786
9	0,9986	0,9982	0,9978	0,9973	0,9967	0,9960	0,9952	0,9942	0,9931	0,9919
10	0,9996	0,9995	0,9994	0,9992	0,9990	0,9987	0,9984	0,9981	0,9977	0,9972
11	0,9999	0,9999	0,9998	0,9998	0,9997	0,9996	0,9995	0,9994	0,9993	0,9991
12	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
13				1,0000	1,0000	1,00000	1,0000	1,0000	0,9999	0,9999

	λ									
k	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9	5,0
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0845	0,0780	0,0719	0,0663	0,0611	0,0563	0,0518	0,0477	0,0439	0,0404
2	0,2238	0,2102	0,1974	0,1851	0,1736	0,1626	0,1523	0,1425	0,1333	0,1247
3	0,4142	0,3954	0,3772	0,3595	0,3423	0,3257	0,3097	0,2942	0,2793	0,2650
4	0,6093	0,5898	0,5704	0,5512	0,5321	0,5132	0,4946	0,4763	0,4582	0,4405
5	0,7693	0,7531	0,7367	0,7199	0,7029	0,6858	0,6684	0,6510	0,6335	0,6160
6	0,8787	0,8675	0,8558	0,8436	0,8311	0,8180	0,8046	0,7908	0,7767	0,7622
7	0,9427	0,9361	0,9290	0,9214	0,9134	0,9050	0,8960	0,8867	0,8769	0,8666
8	0,9755	0,9721	0,9683	0,9642	0,9597	0,9549	0,9497	0,9442	0,9382	0,9319
9	0,9905	0,9889	0,9871	0,9851	0,9829	0,9805	0,9778	0,9749	0,9717	0,9682
10	0,9966	0,9959	0,9952	0,9943	0,9933	0,9922	0,9910	0,9896	0,9880	0,9863
11	0,9989	0,9986	0,9983	0,9980	0,9976	0,9971	0,9966	0,9960	0,9953	0,9945
12	0,9997	0,9996	0,9995	0,9993	0,9992	0,9990	0,9988	0,9986	0,9983	0,9980
13	0,9999	0,9999	0,9998	0,9998	0,9997	0,9997	0,9996	0,9995	0,9994	0,9993
14	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998
15				1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999

	λ									
k	5,1	5,2	5,3	5,4	5,5	5,6	5,7	5,8	5,9	6,0
0	0,0061	0,0055	0,0050	0,0045	0,0041	0,0037	0,0033	0,0030	0,0027	0,0025
1	0,0372	0,0342	0,0314	0,0289	0,0266	0,0244	0,0224	0,0206	0,0189	0,0174
2	0,1165	0,1088	0,1016	0,0948	0,0884	0,0824	0,0768	0,0715	0,0666	0,0620
3	0,2513	0,2381	0,2254	0,2133	0,2017	0,1906	0,1801	0,1700	0,1604	0,1512
4	0,4231	0,4061	0,3895	0,3733	0,3575	0,3422	0,3272	0,3127	0,2987	0,2851
5	0,5984	0,5809	0,5635	0,5461	0,5289	0,5119	0,4950	0,4783	0,4619	0,4457
6	0,7474	0,7324	0,7171	0,7017	0,6860	0,6703	0,6544	0,6384	0,6224	0,6063
7	0,8560	0,8449	0,8335	0,8217	0,8095	0,7970	0,7842	0,7710	0,7576	0,7440
8	0,9552	0,9181	0,9106	0,9027	0,8944	0,8857	0,8766	0,8672	0,8574	0,8472
9	0,9644	0,9603	0,9559	0,9512	0,9462	0,9409	0,9352	0,9292	0,9228	0,9161
10	0,9844	0,9823	0,9800	0,9775	0,9747	0,9718	0,9686	0,9651	0,9614	0,9574
11	0,9937	0,9927	0,9916	0,9904	0,9890	0,9875	0,9859	0,9841	0,9821	0,9799
12	0,9976	0,9972	0,9967	0,9962	0,9955	0,9949	0,9941	0,9932	0,9922	0,9912
13	0,9992	0,9990	0,9988	0,9986	0,9983	0,9980	0,9977	0,9973	0,9969	0,9964
14	0,9997	0,9997	0,9996	0,9995	0,9994	0,9993	0,9991	0,9990	0,9988	0,9986
15	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9996	0,9996	0,9995
16	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998
17				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999

FUNCIÓN DE LA DISTRIBUCIÓN POISSON *(cont.)*

	λ									
k	6,1	6,2	6,3	6,4	6,5	6,6	6,7	6,8	6,9	7,0
0	0,0022	0,0020	0,0018	0,0017	0,0015	0,0014	0,0012	0,0011	0,0010	0,0009
1	0,0159	0,0146	0,0134	0,0123	0,0113	0,0103	0,0095	0,0087	0,0080	0,0073
2	0,0577	0,0536	0,0498	0,0463	0,0430	0,0400	0,0371	0,0344	0,0320	0,0296
3	0,1425	0,1342	0,1264	0,1189	0,1119	0,1052	0,0988	0,0928	0,0871	0,0818
4	0,2719	0,2592	0,2469	0,2351	0,2237	0,2127	0,2022	0,1920	0,1823	0,1730
5	0,4298	0,4141	0,3988	0,3837	0,3690	0,3547	0,3407	0,3270	0,3137	0,3007
6	0,5902	0,5742	0,5582	0,5423	0,5265	0,5108	0,4953	0,4799	0,4647	0,4497
7	0,7301	0,7160	0,7018	0,6873	0,6728	0,6581	0,6433	0,6285	0,6136	0,5987
8	0,8367	0,8259	0,8148	0,8033	0,7916	0,7796	0,7673	0,7548	0,7420	0,7291
9	0,9090	0,9016	0,8939	0,8858	0,8774	0,8686	0,8596	0,8502	0,8405	0,8305
10	0,9531	0,9486	0,9437	0,9386	0,9332	0,9274	0,9214	0,9151	0,9084	0,9015
11	0,9776	0,9750	0,9723	0,9693	0,9661	0,9627	0,9591	0,9552	0,9510	0,9467
12	0,9900	0,9887	0,9873	0,9857	0,9840	0,9821	0,9801	0,9779	0,9755	0,9730
13	0,9958	0,9952	0,9945	0,9937	0,9929	0,9920	0,9909	0,9898	0,9885	0,9872
14	0,9984	0,9981	0,9978	0,9974	0,9970	0,9966	0,9961	0,9956	0,9950	0,9943
15	0,9994	0,9993	0,9992	0,9990	0,9988	0,9986	0,9984	0,9982	0,9979	0,9976
16	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9994	0,9993	0,9992	0,9990
17	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996
18	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
19			1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

	λ									
k	7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8	7,9	8,0
0	0,0008	0,0007	0,0007	0,0006	0,0006	0,0005	0,0005	0,0004	0,0004	0,0003
1	0,0067	0,0061	0,0056	0,0051	0,0047	0,0043	0,0039	0,0036	0,0033	0,0030
2	0,0275	0,0255	0,0236	0,0219	0,0203	0,0188	0,0174	0,0161	0,0149	0,0138
3	0,0767	0,0719	0,0674	0,0632	0,0591	0,0554	0,0518	0,0485	0,0453	0,0424
4	0,1641	0,1555	0,1473	0,1395	0,1321	0,1249	0,1181	0,1117	0,1055	0,0996
5	0,2881	0,2759	0,2640	0,2526	0,2414	0,2307	0,2203	0,2103	0,2006	0,1912
6	0,4349	0,4204	0,4060	0,3920	0,3782	0,3646	0,3514	0,3384	0,3257	0,3134
7	0,5838	0,5689	0,5541	0,5393	0,5246	0,5100	0,4956	0,4812	0,4670	0,4530
8	0,7160	0,7027	0,6892	0,6757	0,6620	0,6482	0,6343	0,6204	0,6065	0,5926
9	0,8202	0,8097	0,7988	0,7877	0,7764	0,7649	0,7531	0,7411	0,7290	0,7166
10	0,8942	0,8867	0,8788	0,8707	0,8622	0,8535	0,8445	0,8352	0,8257	0,8159
11	0,9420	0,9371	0,9319	0,9265	0,9208	0,9148	0,9085	0,9020	0,8952	0,8881
12	0,9703	0,9673	0,9642	0,9609	0,9573	0,9536	0,9496	0,9454	0,9409	0,9362
13	0,9857	0,9841	0,9824	0,9805	0,9784	0,9762	0,9739	0,9714	0,9687	0,9658
14	0,9935	0,9927	0,9918	0,9908	0,9897	0,9886	0,9873	0,9859	0,9844	0,9827
15	0,9972	0,9969	0,9964	0,9959	0,9954	0,9948	0,9941	0,9934	0,9926	0,9918
16	0,9989	0,9987	0,9985	0,9983	0,9980	0,9978	0,9974	0,9971	0,9967	0,9963
17	0,9996	0,9995	0,9994	0,9993	0,9992	0,9991	0,9989	0,9988	0,9986	0,9984
18	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9994	0,9993
19	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997
20	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999
21							1,0000	1,0000	1,0000	1,0000

FUNCIÓN DE LA DISTRIBUCIÓN POISSON *(cont.)*

	λ									
k	8,1	8,2	8,3	8,4	8,5	8,6	8,7	8,8	8,9	9,0
0	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
1	0,0028	0,0025	0,0023	0,0021	0,0019	0,0018	0,0016	0,0015	0,0014	0,0012
2	0,0127	0,0118	0,0109	0,0100	0,0093	0,0086	0,0079	0,0073	0,0068	0,0062
3	0,0396	0,0370	0,0346	0,0323	0,0301	0,0281	0,0262	0,0244	0,0228	0,0212
4	0,0941	0,0887	0,0837	0,0789	0,0744	0,0701	0,0660	0,0621	0,0584	0,0550
5	0,1823	0,1736	0,1653	0,1573	0,1496	0,1422	0,1352	0,1284	0,1219	0,1157
6	0,3013	0,2896	0,2781	0,2670	0,2562	0,2457	0,2355	0,2256	0,2160	0,2068
7	0,4391	0,4254	0,4119	0,3987	0,3856	0,3728	0,3602	0,3478	0,3357	0,3239
8	0,5786	0,5647	0,5508	0,5369	0,5231	0,5094	0,4958	0,4823	0,4689	0,4557
9	0,7041	0,6915	0,6788	0,6659	0,6530	0,6400	0,6269	0,6137	0,6006	0,5874
10	0,8058	0,7956	0,7850	0,7743	0,7634	0,7522	0,7409	0,7294	0,7178	0,7060
11	0,8807	0,8731	0,8652	0,8571	0,8487	0,8400	0,8311	0,8220	0,8126	0,8030
12	0,9313	0,9261	0,9207	0,9150	0,9091	0,9029	0,8965	0,8898	0,8829	0,8758
13	0,9628	0,9595	0,9561	0,9524	0,9486	0,9445	0,9403	0,9358	0,9311	0,9262
14	0,9810	0,9791	0,9771	0,9749	0,9726	0,9701	0,9675	0,9647	0,9617	0,9585
15	0,9908	0,9898	0,9887	0,9875	0,9862	0,9848	0,9832	0,9816	0,9798	0,9780
16	0,9958	0,9953	0,9947	0,9941	0,9934	0,9926	0,9918	0,9909	0,9899	0,9889
17	0,9982	0,9979	0,9977	0,9973	0,9970	0,9966	0,9962	0,9957	0,9952	0,9947
18	0,9992	0,9991	0,9990	0,9989	0,9987	0,9985	0,9983	0,9981	0,9978	0,9976
19	0,9997	0,9997	0,9996	0,9995	0,9995	0,9994	0,9993	0,9992	0,9991	0,9989
20	0,9999	0,9999	0,9998	0,9998	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996
21	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998
22	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999
23	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

	λ									
k	9,1	9,2	9,3	9,4	9,5	9,6	9,7	9,8	9,9	10,0
0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
1	0,0011	0,0010	0,0009	0,0009	0,0008	0,0007	0,0007	0,0006	0,0005	0,0005
2	0,0058	0,0053	0,0049	0,0045	0,0042	0,0038	0,0035	0,0033	0,0030	0,0028
3	0,0198	0,0184	0,0172	0,0160	0,0149	0,0138	0,0129	0,0120	0,0111	0,0103
4	0,0517	0,0486	0,0456	0,0429	0,0403	0,0378	0,0355	0,0333	0,0312	0,0293
5	0,1098	0,1041	0,0987	0,0935	0,0885	0,0838	0,0793	0,0750	0,0710	0,0671
6	0,1978	0,1892	0,1808	0,1727	0,1650	0,1575	0,1502	0,1433	0,1366	0,1301
7	0,3123	0,3010	0,2900	0,2792	0,2687	0,2584	0,2485	0,2388	0,2294	0,2202
8	0,4426	0,4296	0,4168	0,4042	0,3918	0,3796	0,3676	0,3558	0,3442	0,3328
9	0,5742	0,5611	0,5480	0,5349	0,5218	0,5089	0,4960	0,4832	0,4705	0,4579
10	0,6941	0,6820	0,6699	0,6576	0,6453	0,6330	0,6205	0,6081	0,5956	0,5830
11	0,7932	0,7832	0,7730	0,7626	0,7520	0,7412	0,7303	0,7193	0,7081	0,6968
12	0,8684	0,8607	0,8529	0,8448	0,8364	0,8279	0,8191	0,8101	0,8009	0,7916
13	0,9210	0,9156	0,9100	0,9042	0,8981	0,8919	0,8853	0,8786	0,8716	0,8645
14	0,9552	0,9517	0,9480	0,9441	0,9400	0,9357	0,9312	0,9265	0,9216	0,9165
15	0,9760	0,9738	0,9715	0,9691	0,9665	0,9638	0,9609	0,9579	0,9546	0,9513
16	0,9878	0,9865	0,9852	0,9838	0,9823	0,9806	0,9789	0,9770	0,9751	0,9730
17	0,9941	0,9934	0,9927	0,9919	0,9911	0,9902	0,9892	0,9881	0,9870	0,9857
18	0,9973	0,9969	0,9966	0,9962	0,9957	0,9952	0,9947	0,9941	0,9935	0,9928
19	0,9988	0,9986	0,9985	0,9983	0,9980	0,9978	0,9975	0,9972	0,9969	0,9965
20	0,9995	0,9994	0,9993	0,9992	0,9991	0,9990	0,9989	0,9987	0,9986	0,9984
21	0,9998	0,9998	0,9997	0,9997	0,9996	0,9996	0,9995	0,9995	0,9994	0,9993
22	0,9999	0,9999	0,9999	0,9999	0,9999	0,9998	0,9998	0,9998	0,9997	0,9997
23	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
24					1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

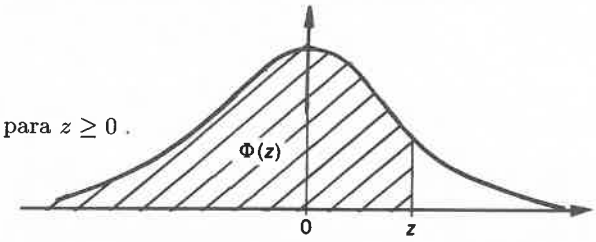
FUNCIÓN DE LA DISTRIBUCIÓN POISSON *(cont.)*

	λ									
k	11,0	12,0	13,0	14,0	15,0	16,0	17,0	18,0	19,0	20,0
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0012	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0049	0,0023	0,0011	0,0005	0,0002	0,0001	0,0000	0,0000	0,0000	0,0000
4	0,0151	0,0076	0,0037	0,0018	0,0009	0,0004	0,0002	0,0001	0,0000	0,0000
5	0,0375	0,0203	0,0107	0,0055	0,0028	0,0014	0,0007	0,0003	0,0002	0,0001
6	0,0786	0,0458	0,0259	0,0142	0,0076	0,0040	0,0021	0,0010	0,0005	0,0003
7	0,1432	0,0895	0,0540	0,0316	0,0180	0,0100	0,0054	0,0029	0,0015	0,0008
8	0,2320	0,1550	0,0998	0,0621	0,0374	0,0220	0,0126	0,0071	0,0039	0,0021
9	0,3405	0,2424	0,1658	0,1094	0,0699	0,0433	0,0261	0,0154	0,0089	0,0050
10	0,4599	0,3472	0,2517	0,1757	0,1185	0,0774	0,0491	0,0304	0,0183	0,0108
11	0,5793	0,4616	0,3532	0,2600	0,1848	0,1270	0,0847	0,0549	0,0347	0,0214
12	0,6887	0,5760	0,4631	0,3585	0,2676	0,1931	0,1350	0,0917	0,0606	0,0390
13	0,7813	0,6815	0,5730	0,4644	0,3632	0,2745	0,2009	0,1426	0,0984	0,0661
14	0,8540	0,7720	0,6751	0,5704	0,4657	0,3675	0,2808	0,2081	0,1497	0,1049
15	0,9074	0,8444	0,7636	0,6694	0,5681	0,4667	0,3715	0,2867	0,2148	0,1565
16	0,9441	0,8987	0,8355	0,7559	0,6641	0,5660	0,4677	0,3751	0,2920	0,2211
17	0,9678	0,9370	0,8905	0,8272	0,7489	0,6593	0,5640	0,4686	0,3784	0,2970
18	0,9823	0,9626	0,9302	0,8826	0,8195	0,7423	0,6550	0,5622	0,4695	0,3814
19	0,9907	0,9787	0,9573	0,9235	0,8752	0,8122	0,7363	0,6509	0,5606	0,4703
20	0,9953	0,9884	0,9750	0,9521	0,9170	0,8682	0,8055	0,7307	0,6472	0,5591
21	0,9977	0,9939	0,9859	0,9712	0,9469	0,9108	0,8615	0,7991	0,7255	0,6437
22	0,9990	0,9970	0,9924	0,9833	0,9673	0,9418	0,9047	0,8551	0,7931	0,7206
23	0,9995	0,9985	0,9960	0,9907	0,9805	0,9633	0,9367	0,8989	0,8490	0,7875
24	0,9998	0,9993	0,9980	0,9950	0,9888	0,9777	0,9594	0,9317	0,8933	0,8432
25	0,9999	0,9997	0,9990	0,9974	0,9938	0,9869	0,9748	0,9554	0,9269	0,8878
26	1,0000	0,9999	0,9995	0,9987	0,9967	0,9925	0,9848	0,9718	0,9514	0,9221
27		0,9999	0,9998	0,9994	0,9983	0,9959	0,9912	0,9827	0,9687	0,9475
28		1,0000	0,9999	0,9997	0,9991	0,9978	0,9950	0,9897	0,9805	0,9657
29			1,0000	0,9999	0,9996	0,9989	0,9973	0,9941	0,9882	0,9782
30			1,0000	0,9999	0,9998	0,9994	0,9986	0,9967	0,9930	0,9865
31				1,0000	0,9999	0,9997	0,9993	0,9982	0,9960	0,9919
32					1,0000	0,9999	0,9996	0,9990	0,9978	0,9953
33						0,9999	0,9998	0,9995	0,9988	0,9973
34						1,0000	0,9999	0,9998	0,9994	0,9985
35							1,0000	0,9999	0,9997	0,9992
36								0,9999	0,9998	0,9996
37								1,0000	0,9999	0,9998
38									1,0000	0,9999

III. Función de distribución normal tipificada

La función tabulada es:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx \qquad \text{para } z \geq 0 .$$



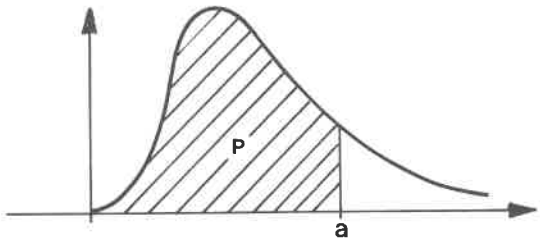
Si $z < 0$, $\Phi(z) = 1 - \Phi(-z)$ por ser la función de densidad simétrica respecto del cero.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,34448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	$0,9^20097$	$0,9^20358$	$0,9^20613$	$0,9^20863$	$0,9^21106$	$0,9^21344$	$0,9^21576$
2,4	$0,9^21802$	$0,9^22024$	$0,9^22240$	$0,9^22451$	$0,9^22656$	$0,9^22857$	$0,9^23053$	$0,9^23244$	$0,9^23431$	$0,9^23613$
2,5	$0,9^23790$	$0,9^23963$	$0,9^24132$	$0,9^24297$	$0,9^24457$	$0,9^24614$	$0,9^24766$	$0,9^24915$	$0,9^25060$	$0,9^25201$
2,6	$0,9^25339$	$0,9^25473$	$0,9^25604$	$0,9^25731$	$0,9^25855$	$0,9^25975$	$0,9^26093$	$0,9^26207$	$0,9^26319$	$0,9^26427$
2,7	$0,9^26533$	$0,9^26636$	$0,9^26736$	$0,9^26833$	$0,9^26928$	$0,9^27020$	$0,9^27110$	$0,9^27197$	$0,9^27282$	$0,9^27365$
2,8	$0,9^27445$	$0,9^27523$	$0,9^27599$	$0,9^27673$	$0,9^27744$	$0,9^27814$	$0,9^27882$	$0,9^27948$	$0,9^28012$	$0,9^28074$
2,9	$0,9^28134$	$0,9^28193$	$0,9^28250$	$0,9^28305$	$0,9^28359$	$0,9^28411$	$0,9^28462$	$0,9^28511$	$0,9^28559$	$0,9^28605$
3,0	$0,9^28650$	$0,9^28694$	$0,9^28736$	$0,9^28777$	$0,9^28817$	$0,9^28856$	$0,9^28893$	$0,9^28930$	$0,9^28965$	$0,9^28999$
3,1	$0,9^30324$	$0,9^30646$	$0,9^30957$	$0,9^31260$	$0,9^31553$	$0,9^31836$	$0,9^32112$	$0,9^32378$	$0,9^32636$	$0,9^32886$
3,2	$0,9^33129$	$0,9^33363$	$0,9^33590$	$0,9^33810$	$0,9^34024$	$0,9^34230$	$0,9^34429$	$0,9^34623$	$0,9^34810$	$0,9^34991$
3,3	$0,9^35166$	$0,9^35335$	$0,9^35499$	$0,9^35658$	$0,9^35811$	$0,9^35959$	$0,9^36103$	$0,9^36242$	$0,9^36376$	$0,9^36505$
3,4	$0,9^36631$	$0,9^36752$	$0,9^36869$	$0,9^36982$	$0,9^37091$	$0,9^37197$	$0,9^37299$	$0,9^37398$	$0,9^37493$	$0,9^37585$
3,5	$0,9^37674$	$0,9^37759$	$0,9^37842$	$0,9^37922$	$0,9^37999$	$0,9^38074$	$0,9^38146$	$0,9^38215$	$0,9^38282$	$0,9^38347$
3,6	$0,9^38409$	$0,9^38469$	$0,9^38527$	$0,9^38583$	$0,9^38637$	$0,9^38689$	$0,9^38739$	$0,9^38787$	$0,9^38834$	$0,9^38879$
3,7	$0,9^38922$	$0,9^38964$	$0,9^40039$	$0,9^40426$	$0,9^40799$	$0,9^41158$	$0,9^41504$	$0,9^41838$	$0,9^42159$	$0,9^42468$
3,8	$0,9^42765$	$0,9^43052$	$0,9^43327$	$0,9^43593$	$0,9^43848$	$0,9^44094$	$0,9^44331$	$0,9^44558$	$0,9^44777$	$0,9^44988$
3,9	$0,9^45190$	$0,9^45385$	$0,9^45573$	$0,9^45753$	$0,9^45926$	$0,9^46092$	$0,9^46253$	$0,9^46406$	$0,9^46554$	$0,9^46696$
4,0	$0,9^46833$	$0,9^46964$	$0,9^47090$	$0,9^47211$	$0,9^47327$	$0,9^47439$	$0,9^47546$	$0,9^47649$	$0,9^47748$	$0,9^47843$

IV. Función de distribución χ^2 de Pearson

La tabla proporciona el valor a, tal que $p[X \leq a] = p$, donde X es una variable con distribución χ^2 de Pearson con n grados de libertad ($n \leq 100$).

Para $n > 100$, la distribución de $\sqrt{2 \cdot X}$ se aproxima a una $N(\sqrt{2n-1}, 1)$.



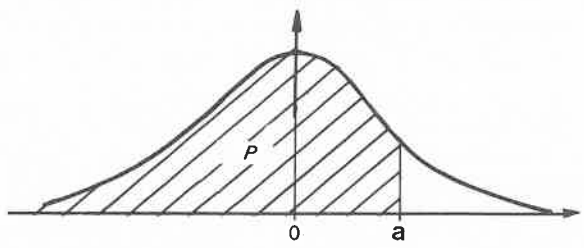
n \ p	0,995	0,99	0,975	0,95	0,90	0,75	0,50	0,25	0,10	0,05	0,025	0,01	0,005
1	7,88	6,63	5,02	3,84	2,71	1,32	0,455	0,102	0,0158	0,0039	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,77	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,63	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,989
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28,3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34,3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	30,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67.0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	116,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

V. Función de distribución t de Student

La tabla proporciona el valor a, tal que $p[X \leq a] = p$, para $p > 0.5$, donde X es una variable con distribución t de Student con n grados de libertad ($n = 1, \ldots, 30, \infty$).

Para $p \leq 0.5$, $p[X \leq a] = 1 - p[X \leq -a]$, al ser la función de densidad simétrica respecto del cero.

Para valores grandes de n ($n > 30$), la distribución t se aproxima a la normal tipificada.



n \ p	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	0,975	0,99	0,995	0,9995
1	0,158	0,325	0,510	0,727	1,00	1,38	1,96	3,08	6,31	12,7	31,8	63,7	637
2	0,142	0,289	0,445	0,617	0,816	1,06	1,39	1,89	2,92	4,30	6,96	9,92	31,6
3	0,137	0,277	0,424	0,584	0,765	0,978	1,25	1,64	2,35	3,18	4,54	5,84	12,9
4	0,134	0,271	0,414	0,569	0,741	0,941	1,19	1,53	2,13	2,78	3,75	4,60	8,61
5	0,132	0,267	0,408	0,559	0,727	0,920	1,16	1,48	2,01	2,57	3,36	4,03	6,86
6	0,131	0,265	0,404	0,553	0,718	0,906	1,13	1,44	1,94	2,45	3,14	3,71	5,96
7	0,130	0,263	0,402	0,549	0,711	0,896	1,12	1,42	1,90	2,36	3,00	3,50	5,40
8	0,130	0,262	0,399	0,546	0,706	0,889	1,11	1,40	1,86	2,31	2,90	3,36	5,04
9	0,129	0,261	0,398	0,543	0,703	0,883	1,10	1,38	1,83	2,26	2,82	3,25	4,78
10	0,129	0,260	0,397	0,542	0,700	0,879	1,09	1,37	1,81	2,23	2,76	3,17	4,59
11	0,129	0,260	0,396	0,540	0,697	0,876	1,09	1,36	1,80	2,20	2,72	3,11	4,44
12	0,128	0,259	0,395	0,539	0,695	0,873	1,08	1,36	1,78	2,18	2,68	3,06	4,32
13	0,128	0,259	0,394	0,538	0,694	0,870	1,08	1,35	1,77	2,16	2,65	3,01	4,22
14	0,128	0,258	0,393	0,537	0,692	0,868	1,08	1,34	1,76	2,14	2,62	2,98	4,14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,07	1,34	1,75	2,13	2,60	2,95	4,07
16	0,128	0,258	0,392	0,535	0,690	0,865	1,07	1,34	1,75	2,12	2,58	2,92	4,02
17	0,128	0,257	0,392	0,534	0,689	0,863	1,07	1,33	1,74	2,11	2,57	2,90	3,96
18	0,127	0,257	0,392	0,534	0,688	0,862	1,07	1,33	1,73	2,10	2,55	2,88	3,92
19	0,127	0,257	0,391	0,533	0,688	0,861	1,07	1,33	1,73	2,09	2,54	2,86	3,88
20	0,127	0,257	0,391	0,533	0,687	0,860	1,06	1,32	1,72	2,09	2,53	2,84	3,85
21	0,127	0,257	0,391	0,532	0,686	0,859	1,06	1,32	1,72	2,08	2,52	2,83	3,82
22	0,127	0,256	0,390	0,532	0,686	0,858	1,06	1,32	1,72	2,07	2,51	2,82	3,79
23	0,127	0,256	0,390	0,532	0,685	0,858	1,06	1,32	1,71	2,07	2,50	2,81	3,77
24	0,127	0,256	0,390	0,531	0,685	0,857	1,06	1,32	1,71	2,06	2,49	2,80	3,74
25	0,127	0,256	0,390	0,531	0,684	0,856	1,06	1,32	1,71	2,06	2,48	2,79	3,72
26	0,127	0,256	0,390	0,531	0,684	0,856	1,06	1,32	1,70	2,06	2,48	2,78	3,71
27	0,127	0,256	0,389	0,531	0,684	0,855	1,06	1,31	1,70	2,05	2,47	2,77	3,69
28	0,127	0,256	0,389	0,530	0,683	0,855	1,06	1,31	1,70	2,05	2,47	2,76	3,67
29	0,127	0,256	0,389	0,530	0,683	0,854	1,05	1,31	1,70	2,04	2,46	2,76	3,66
30	0,127	0,256	0,389	0,530	0,683	0,854	1,05	1,31	1,70	2,04	2,46	2,75	3,65
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,04	1,28	1,64	1,96	2,33	2,58	3,29

VI. Función de distribución F de Snedecor

Las tablas proporcionan el valor a, tal que $p[X \leq a] = p$, para $p = 0.90$, $p = 0.95$ y $p = 0.99$, donde X es una variable con distribución F de Snedecor con n_1 y n_2 grados de libertad $(n_1, n_2 = 1, \ldots, 120, \infty)$.

De la génesis de la distribución F de Snedecor se deduce:

$$p\left[F_{n_1,n_2} \leq a\right] = 1 - p\left[F_{n_2,n_1} \leq \frac{1}{a}\right]$$

$$\boxed{p[X \leq a] = 0.90}$$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39,864	49,500	53,593	55,833	57,241	58,204	58,906	59,439	59,858	60,195	60,705	61,220	61,740	62,002	62,265	62,529	62,794	63,061	63,328
2	8,5263	9,0000	9,1618	9,2434	9,2926	9,3255	9,3491	9,3668	9,3805	9,3916	9,4081	9,4247	9,4413	9,4496	9,4579	9,4663	9,4746	9,4829	9,4913
3	5,5383	5,4624	5,3908	5,3427	5,3092	5,2847	5,2662	5,2517	5,2400	5,2304	5,2156	5,2003	5,1845	5,1764	5,1681	5,1597	5,1512	5,1425	5,1337
4	4,5448	4,3246	4,1908	4,1073	4,0506	4,0098	3,9790	3,9549	3,9357	3,9199	3,8955	3,8689	3,8443	3,8310	3,8174	3,8036	3,7896	3,7753	3,7607
5	4,0604	3,7797	3,6195	3,5202	3,4530	3,4045	3,3679	3,3393	3,3163	3,2974	3,2682	3,2380	3,2067	3,1905	3,1741	3,1573	3,1402	3,1228	3,1050
6	3,7760	3,4633	3,2888	3,1808	3,1075	3,0540	3,0145	2,9830	2,9577	2,9369	2,9047	2,8712	2,8363	2,8183	2,8000	2,7812	2,7620	2,7423	2,7222
7	3,5894	3,2574	3,0741	2,9605	2,8833	2,8274	2,7849	2,7516	2,7247	2,7025	2,6681	2,6322	2,5947	2,5753	2,5555	2,5351	2,5142	2,4928	2,4708
8	3,4579	3,1131	2,9238	2,8064	2,7265	2,6683	2,6241	2,5893	2,5612	2,5380	2,5020	2,4642	2,4246	2,4041	2,3830	2,3614	2,3391	2,3162	2,2926
9	3,3603	3,0065	2,8129	2,6927	2,6106	2,5509	2,5053	2,4694	2,4403	2,4163	2,3789	2,3396	2,2983	2,2768	2,2547	2,2320	2,2085	2,1843	2,1592
10	3,2850	2,9245	2,7277	2,6053	2,5216	2,4606	2,4140	2,3772	2,3473	2,3226	2,2841	2,2435	2,2007	2,1784	2,1554	2,1317	2,1072	2,0818	2,0554
11	3,2252	2,8595	2,6602	2,5362	2,4512	2,3891	2,3416	2,3040	2,2735	2,2482	2,2087	2,1671	2,1230	2,1000	2,0762	2,0516	2,0261	1,9997	1,9721
12	3,1765	2,8068	2,6055	2,4801	2,3940	2,3310	2,2828	2,2446	2,2135	2,1878	2,1474	2,1049	2,0597	2,0360	2,0115	1,9861	1,9597	1,9323	1,9036
13	3,1362	2,7632	2,5603	2,4337	2,3467	2,2830	2,2341	2,1953	2,1638	2,1376	2,0966	2,0532	2,0070	1,9827	1,9576	1,9315	1,9043	1,8759	1,8462
14	3,1022	2,7265	2,5222	2,3947	2,3069	2,2426	2,1931	2,1539	2,1220	2,0954	2,0537	2,0095	1,9625	1,9377	1,9119	1,8852	1,8572	1,8280	1,7973
15	3,0732	2,6952	2,4898	2,3614	2,2730	2,2081	2,1582	2,1185	2,0862	2,0593	2,0171	1,9722	1,9243	1,8990	1,8728	1,8454	1,8168	1,7867	1,7551
16	3,0481	2,6682	2,4618	2,3327	2,2438	2,1783	2,1280	2,0880	2,0553	2,0281	1,9854	1,9399	1,8913	1,8656	1,8388	1,8108	1,7816	1,7507	1,7182
17	3,0262	2,6446	2,4374	2,3077	2,2183	2,1524	2,1017	2,0613	2,0284	2,0009	1,9577	1,9117	1,8624	1,8362	1,8090	1,7805	1,7506	1,7191	1,6856
18	3,0070	2,6239	2,4160	2,5858	2,1958	2,1296	2,0785	2,0379	2,0047	1,9770	1,9333	1,8868	1,8368	1,8103	1,7827	1,7537	1,7232	1,6910	1,6567
19	2,9899	2,6056	2,3970	2,2663	2,1760	2,1094	2,0580	2,0171	1,9836	1,9557	1,9117	1,8647	1,8142	1,7667	1,7382	1,7083	1,6988	1,6659	1,6308
20	2,9747	2,5893	2,3801	2,2489	2,1582	2,0913	2,0397	1,9985	1,9649	1,9367	1,8924	1,8449	1,7938	1,7873	1,7592	1,7298	1,6768	1,6433	1,6074
21	2,9609	2,5746	2,3649	2,2333	2,1423	2,0751	2,0232	1,9819	1,9480	1,9197	1,8750	1,8272	1,7756	1,7481	1,7193	1,6890	1,6569	1,6228	1,5862
22	2,9486	2,5613	2,3512	2,2193	2,1279	2,0605	2,0084	1,9668	1,9327	1,9043	1,8593	1,8111	1,7590	1,7312	1,7021	1,6714	1,6389	1,6042	1,5668
23	2,9374	2,5493	2,3387	2,2065	2,1149	2,0472	1,9949	1,9531	1,9189	1,8903	1,8450	1,7964	1,7439	1,7159	1,6864	1,6554	1,6224	1,5871	1,5490
24	2,9271	2,5383	2,3274	2,1949	2,1030	2,0351	1,9826	1,9407	1,9063	1,8775	1,8319	1,7831	1,7302	1,7019	1,6721	1,6407	1,6073	1,5715	1,5327
25	2,9177	2,5283	2,3170	2,1843	2,0922	2,0241	1,9714	1,9292	1,8947	1,8658	1,8200	1,7708	1,7175	1,6890	1,6589	1,6272	1,5934	1,5570	1,5176
26	2,9091	2,5191	2,3075	2,1745	2,0822	2,0139	1,9610	1,9188	1,8841	1,8550	1,8090	1,7596	1,7059	1,6771	1,6468	1,6147	1,5805	1,5437	1,5036
27	2,9012	2,5106	2,2987	2,1655	2,0730	2,0045	1,9515	1,9091	1,8743	1,8451	1,7989	1,7492	1,6951	1,6662	1,6356	1,6032	1,5686	1,5313	1,4906
28	2,8939	2,5028	2,2906	2,1571	2,0645	1,9959	1,9427	1,9001	1,8652	1,8359	1,7895	1,7395	1,6852	1,6560	1,6252	1,5925	1,5575	1,5198	1,4784
29	2,8871	2,4955	2,2831	2,1494	2,0566	1,9878	1,9345	1,8918	1,8560	1,8274	1,7808	1,7306	1,6759	1,6465	1,6155	1,5825	1,5472	1,5090	1,4670
30	2,8807	2,4887	2,2761	2,1422	2,0492	1,9803	1,9269	1,8841	1,8498	1,8195	1,7727	1,7223	1,6673	1,6377	1,6065	1,5732	1,5376	1,4989	1,4564
40	2,8354	2,4404	2,2261	2,0909	1,9968	1,9269	1,8725	1,8289	1,7929	1,7627	1,7146	1,6624	1,6052	1,5741	1,5411	1,5056	1,4672	1,4248	1,3769
60	2,7914	2,3932	2,1774	2,0410	1,9457	1,8747	1,8194	1,7748	1,7380	1,7070	1,6574	1,6034	1,5435	1,5107	1,4755	1,4373	1,3952	1,3476	1,2915
120	2,7478	2,3473	2,1300	1,9923	1,8959	1,8238	1,7675	1,7220	1,6843	1,6524	1,6012	1,5450	1,4821	1,4472	1,4094	1,3676	1,3203	1,2646	1,1926
∞	2,7055	2,3026	2,0838	1,9449	1,8473	1,7741	1,7167	1,6702	1,6315	1,5987	1,5458	1,4871	1,4206	1,3832	1,3419	1,2951	1,2400	1,1686	1,0000

PERCENTILES DE ORDEN 0,95 DE LA DISTRIBUCIÓN F con n_1 grados de libertad en el numerador y n_2 en el denominador

$$p[X \leq a] = 0.95$$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	245,95	248,01	249,05	250,09	251,14	252,20	253,25	254,32
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,8452	8,8123	8,7855	8,7446	8,7029	8,6602	8,6385	8,6166	8,5944	8,5720	8,5494	8,5265
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	6,0410	5,9988	5,9644	5,9117	5,8578	5,8025	5,7744	5,7459	5,7170	5,6878	5,6581	5,6281
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725	4,7351	4,6777	4,6188	4,5581	4,5272	4,4957	4,4638	4,4314	4,3984	4,3650
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,1468	4,0990	4,0600	3,9999	3,9381	3,8742	3,8415	3,8082	3,7743	3,7398	3,7047	3,6688
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767	3,6365	3,5747	3,5108	3,4445	3,4105	3,3758	3,3404	3,3043	3,2674	3,2298
8	5,3177	4,4590	4,0662	3,8378	3,6875	3,5806	3,5005	3,4281	3,3881	3,3472	3,2840	3,2184	3,1503	3,1152	3,0794	3,0428	3,0053	2,9669	2,9276
9	5,1174	4,2565	3,8626	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789	3,1373	3,0729	3,0061	2,9365	2,9005	2,8637	2,8259	2,7872	2,7475	2,7067
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204	2,9782	2,9130	2,8450	2,7740	2,7372	2,6996	2,6609	2,6211	2,5801	2,5379
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962	2,8536	2,7876	2,7186	2,6464	2,6090	2,5705	2,5309	2,4901	2,4480	2,4045
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964	2,7534	2,6866	2,6169	2,5436	2,5055	2,4663	2,4259	2,3842	2,3410	2,2962
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144	2,6710	2,6037	2,5331	2,4589	2,4202	2,3803	2,3392	2,2966	2,2524	2,2064
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458	2,6021	2,5342	2,4630	2,3879	2,3487	2,3082	2,2664	,2230	2,1778	2,1307
15	4,5431	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066	2,6408	2,5876	2,5437	2,4753	2,4035	2,3275	2,2878	2,2468	2,2043	2,1601	2,1141	2,0658
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377	2,4935	2,4247	2,3522	2,2756	2,2354	2,1938	2,1507	2,1058	2,0589	2,0096
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943	2,4499	2,3807	2,3077	2,2304	2,1898	2,1477	2,1040	2,0584	2,0107	1,9604
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563	2,4117	2,3421	2,2686	2,1906	2,1497	2,1071	2,0629	2,0166	1,9681	1,9168
19	4,3808	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227	2,3779	2,3080	2,2341	2,1555	2,1141	2,0712	2,0264	1,9796	1,9302	1,8780
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5900	2,5140	2,4471	2,3928	2,3479	2,2776	2,2033	2,1242	2,0825	2,0391	1,9938	1,9464	1,8963	1,8432
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3661	2,3210	2,2504	2,1757	2,0960	2,0540	2,0102	1,9645	1,9165	1,8657	1,8117
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419	2,2967	2,2258	2,1508	2,0707	2,0283	1,9842	1,9380	1,8895	1,8380	1,7831
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201	2,2747	2,2036	2,1882	2,0476	2,0050	1,9605	1,9139	1,8649	1,8128	1,7570
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002	2,2547	2,1834	2,1077	2,0267	1,9838	1,9390	1,8920	1,8424	1,7897	1,7331
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821	2,3365	2,1649	2,0889	2,0075	1,9643	1,9192	1,8718	1,8217	1,7684	1,7110
26	4,2252	3,3690	2,9751	2,7426	2,5868	2,4741	2,3883	2,3205	2,2655	2,2197	2,1479	2,0716	1,9898	1,9464	1,9010	1,8533	1,8027	1,7488	1,6906
27	4,2100	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732	2,3053	2,2501	2,2043	2,1323	2,0558	1,9736	1,9299	1,8842	1,8361	1,7851	1,7307	1,6717
28	4,1960	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593	2,2913	2,2360	2,1900	2,1179	2,0411	1,9586	1,9147	1,8687	1,8203	1,7689	1,7138	1,6541
29	4,1830	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463	2,2782	2,2229	2,1768	2,1045	2,0275	1,9446	1,9005	1,8543	1,8055	1,7537	1,6981	1,6377
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107	2,1646	2,0921	2,0148	1,9317	1,8874	1,8409	1,7918	1,7396	1,6835	1,6223
40	4,0848	3,2317	2,8387	2,6060	2,4495	2,3359	2,2490	2,1802	2,1240	2,0772	2,0035	1,9245	1,8389	1,6929	1,7444	1,6928	1,6373	1,5766	1,5089
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2540	2,1665	2,0970	2,0401	1,9926	1,9174	1,8364	1,7480	1,7001	1,6491	1,5943	1,5343	1,4673	1,3893
120	3,9201	3,0718	2,6802	2,4472	2,2900	2,1750	2,0867	2,0164	1,9588	1,9105	1,8337	1,7505	1,6587	1,6084	1,5543	1,4952	1,4290	1,3519	1,2539
∞	2,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799	1,8307	1,7522	1,6664	1,5705	1,5173	1,4591	1,3940	1,3180	1,2214	1,0000

PERCENTILES DE ORDEN 0,99 DE LA DISTRIBUCIÓN F con n_1 grados de libertad en el numerador y n_2 en el denominador

$$p[X \leq a] = 0.99$$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5	6055,8	6106,3	6157,3	6708,7	6234,6	6260,7	6286,8	6313,0	6339,4	6366,0
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388	99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	15,558	13,463
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158	10,051	9,8883	9,7222	9,5527	9,4665	9,3793	9,2912	9,2020	9,1118	9,0204
6	13,745	10,925	9,7795	9,1483	8,7459	8,4661	8,2600	8,1016	7,9761	7,8741	7,7183	7,5590	7,3958	7,3127	7,2285	7,1432	7,0568	6,9690	6,8801
7	12,246	9,5466	8,4513	7,8467	7,4604	7,1914	6,9928	6,8401	6,7188	6,6201	6,4691	6,3143	6,1554	6,0743	5,9921	5,9084	5,8236	5,7372	5,6495
8	11,259	8,6491	7,5910	7,0060	6,6318	6,3707	6,1776	6,0289	5,9106	5,8143	5,6668	5,5151	5,3591	5,2793	5,1980	5,1156	5,0316	4,9460	4,8588
9	10,561	8,0215	6,9919	6,4221	6,0569	5,8018	5,6129	5,4071	5,3511	5,2565	5,1114	4,9621	4,8080	4,7290	4,6486	4,5667	4,4831	4,3978	4,3105
10	10,044	7,5194	6,5523	5,9943	5,6363	5,3858	5,2001	5,0567	4,9424	4,0492	4,7059	4,5582	4,4054	4,3269	4,2469	4,1653	4,0819	3,9965	3,9090
11	9,6400	7,2057	6,2167	5,6683	5,3160	5,0692	4,8861	4,7445	4,6315	4,5393	4,3974	4,2509	4,0990	4,0209	3,9411	3,8596	3,7761	3,6904	3,6025
12	9,3302	6,9266	5,9526	5,4119	5,0643	4,8206	4,6395	4,4994	4,3875	4,2961	4,1553	4,0096	3,8584	3,7805	3,7008	3,6192	3,5355	3,4494	3,3608
13	9,0708	6,7010	5,7394	5,2053	4,8616	4,6204	4,4410	4,3021	4,1911	4,1003	3,9603	3,8154	3,6646	3,5868	3,5070	3,4253	3,3413	3,2548	3,1654
14	8,8616	6,5149	5,5639	5,0354	4,6950	4,4558	4,2779	4,1399	4,0297	3,9394	3,8001	3,6557	3,5052	3,4274	3,3476	3,2656	3,1813	3,0942	3,0040
15	8,6831	6,3589	5,4170	4,8932	4,5556	4,3183	4,1415	4,0045	3,8948	3,8049	3,6662	3,5222	3,3719	3,2940	3,2141	3,1319	3,0471	2,9595	2,8684
16	8,5310	6,2262	5,2922	4,7726	4,4374	4,2016	4,2059	3,8896	3,7804	3,6909	3,5527	3,4089	3,2588	3,1808	3,1007	3,0182	2,9330	2,8447	2,7528
17	8,3997	6,1121	5,1850	4,6690	4,3359	4,1015	3,9267	3,7910	3,6822	3,5931	3,4552	3,3117	3,1615	3,0835	3,0032	2,9205	2,8348	2,7459	2,6530
18	8,2854	6,0129	5,0919	4,5790	4,2479	4,0146	3,8406	3,7054	3,5971	3,5082	3,3706	3,2273	3,0771	2,9990	2,9185	2,8354	2,7493	2,6597	2,5660
19	8,1850	5,9259	5,0103	4,5003	4,1704	3,9386	3,7653	3,6305	3,5225	3,4338	3,2965	3,1533	3,0031	2,9249	2,8442	2,7608	2,6742	2,5839	2,4893
20	8,0960	5,8489	5,9382	4,4307	4,1027	3,8714	3,6987	3,5644	3,4567	3,3682	3,2311	3,0880	2,9377	2,8594	2,7785	2,6947	2,6077	2,5168	2,4212
21	8,0166	5,7804	4,8740	4,3688	4,0421	3,8117	3,6396	3,5056	3,3981	3,3098	3,1729	3,0299	2,8796	2,8011	2,7200	2,6359	2,5484	2,4568	2,3603
22	7,9454	5,7190	4,8166	4,3134	3,9880	3,7583	3,5867	3,4530	3,3458	3,2576	3,1209	2,9780	2,8274	2,7488	2,6675	2,5831	2,4951	2,4029	2,3055
23	7,8811	5,6637	4,7649	4,2635	3,9392	3,7102	3,5390	3,4057	3,2986	3,2106	3,0740	2,9311	2,7805	2,7017	2,6202	2,5355	2,4471	2,3542	2,2559
24	7,8229	5,6136	4,7181	4,2184	3,8951	3,6667	3,4959	3,3679	3,2560	3,1681	3,0316	2,8887	2,7380	2,6591	2,5773	2,4923	2,4035	2,3099	2,2107
25	7,7698	5,5680	4,6755	4,1774	3,8550	3,6272	3,4568	3,3239	3,2172	3,1294	2,9931	2,8502	2,6993	2,6203	2,5383	2,4530	2,3637	2,2695	2,1694
26	7,7213	5,5263	4,6166	4,1400	3,8183	3,5911	3,4210	3,2884	3,1818	3,0941	2,9579	2,8150	2,6640	2,5848	2,5026	2,4170	2,3273	2,2325	2,1315
27	7,6767	5,4881	4,0009	4,1056	3,7848	3,5580	3,3882	3,2558	3,1494	3,0618	2,9256	2,7827	2,6316	2,5522	2,4699	2,3840	2,2938	2,1984	2,0965
28	7,6356	5,4529	4,5681	4,0740	3,7539	3,5276	3,3581	3,2259	3,1195	3,0320	2,8959	2,7530	2,6017	2,5223	2,4397	2,3535	2,2629	2,1670	2,0642
29	7,5976	5,4205	4,5378	4,0449	3,7254	3,4995	3,3302	3,1982	3,0920	3,0045	2,8685	2,7256	2,5742	2,4946	2,4118	2,3253	2,2344	2,1378	2,0342
30	7,5625	5,3904	4,5097	4,0179	3,6990	3,4735	3,3045	3,1726	3,0665	2,9791	2,8431	2,7002	2,5487	2,4689	2,3860	2,2992	2,2079	2,1107	2,0062
40	7,3141	5,1785	4,3126	3,8283	3,5138	3,2910	3,1238	2,9930	2,8876	2,8005	2,6648	2,5216	2,3689	2,2880	2,2034	2,1162	2,0194	1,9172	1,8047
60	7,0771	4,9774	4,1259	3,6491	3,3389	3,1187	2,9530	2,8233	2,7185	2,6318	2,4961	2,3523	2,1978	2,1154	2,0285	1,9360	1,8363	1,7263	1,6006
120	6,8510	4,7865	3,9493	3,4796	3,1735	2,9559	2,7918	2,6629	2,5586	2,4721	2,3363	2,1915	2,0346	1,9500	1,8600	1,7628	1,0557	1,5530	1,3805
∞	6,6349	4,6052	3,7816	3,3192	3,0173	2,8020	2,6393	2,5113	2,4073	2,3209	2,1848	2,0385	1,8783	1,7908	1,6964	1,5923	1,4730	1,3246	1,0000

VII. Tabla de números aleatorios

La tabla proporciona 2500 dígitos aleatorios, ordenados en bloques de cinco.

49914	68221	56088	06112	95481	30094	68493	88796	94771	89418
62045	40681	15941	05962	44378	64349	29782	33287	27400	42915
42534	31925	94158	90197	62874	53659	33433	48610	14698	54761
76126	41049	43363	52461	00552	93352	58497	16347	87145	73668
80434	73037	69008	36801	25520	14161	32300	04187	80668	07499
81301	39731	53857	19690	39998	49829	12399	70867	44498	17385
54521	42350	82908	51212	70208	39891	64871	67448	42988	32600
82530	22869	87826	06678	36873	61198	87748	07531	29592	39612
81338	64309	45798	42954	95565	02789	83017	82936	67117	17709
58264	60374	32610	17879	96900	68029	06993	84288	35401	56317
77023	46829	21332	77383	15547	29332	77698	89878	20489	71800
29750	59902	78110	59018	87548	10225	15774	70778	56086	08117
08288	38411	69886	64918	29055	87607	37452	38174	31431	46173
93908	94810	22057	94240	89918	16561	92716	66461	22337	64718
06341	25883	42574	80202	57287	95120	69332	19036	43326	98697
23240	94741	55622	79479	34606	51079	09476	10695	49618	63037
96370	19171	40441	05002	33165	28693	45027	73791	23047	32976
97050	16194	61095	26533	81738	77032	60551	31605	95212	81078
40833	12169	10712	78345	48236	45086	61654	94929	69169	70561
95676	13582	25664	60838	88071	50052	63188	50346	65618	17517
28030	14185	13226	99566	45483	10079	22945	23903	11695	10694
60202	32586	87466	83357	95516	31258	66309	40615	30572	60842
46530	48755	02308	79508	53422	50805	08896	06963	93922	99423
53151	95839	01745	46462	81463	28669	60179	17880	75875	34562
80272	64398	88249	06792	98424	66842	49129	98939	34173	49883
87024	74221	69721	44518	58804	04860	18127	16855	61558	15430
04852	03436	72753	99836	37513	91341	53517	92094	54386	44563
33592	45845	52015	72030	23071	92933	84219	39455	57792	14216
68121	53688	56812	34869	28573	51079	94677	23993	88241	97735
25062	10428	43930	69033	73395	83469	25990	12971	73728	03856
78183	44396	11064	92153	96293	00825	21079	78337	19739	13684
70209	23316	32828	00927	61841	64754	91125	01206	06691	50868
94342	91040	94035	02650	36284	91162	07950	36178	42536	49869
92503	29854	24116	61149	49266	82303	54924	58251	23928	20703
71646	57503	82416	22657	72359	30085	13037	39608	77439	49318
51809	70780	41544	27828	84321	07714	25865	97896	01924	62028
88504	21620	07292	71021	80929	45042	08703	45894	24521	49942
33186	49273	87542	41086	29615	81101	43707	87031	36101	15137
40068	35043	05280	62921	30122	65119	40512	26855	40842	83244
76401	68461	20711	12007	19209	28259	49820	76415	51534	63574
47014	93729	74235	47808	52473	03145	92563	05837	70023	33169
67147	48017	90741	53647	55007	36607	29360	83163	79024	26155
86987	62924	93157	70947	07336	49541	81386	26968	38311	99885
58973	47026	78574	08804	22960	32850	67944	92303	61216	72948
71635	86749	40369	94639	40731	54012	03972	98581	45604	34885
60971	54212	32596	03052	84150	36798	62635	26210	95685	87089
06599	60910	66315	96690	19039	39878	44688	65146	02482	73130
89960	27162	66264	71024	18708	77974	40473	87155	35834	03114
03930	56898	61900	44036	90012	17673	54167	82396	39468	49566
31338	28729	02095	07429	35718	86882	37513	51560	08872	33717

GLOSARIO

ACHENWALL, G., 3, 13
ADRIAN, R., 493
Aleatorio, experimento, 5, 21
—, fenómeno, 21
ALEKSANDROV, A. D., 9
ARISTÓTELES, 13
Asintótica, distribución, 523
Átomos de una σ-álgebra, 30

$b(p)$ (*distribución de Bernoulli*), 392
$\mathcal{B}(n,p)$ (*distribución binomial*), 394
$\mathcal{BN}(r,p)$ (*distribución binomial negativa*), 407
BAYES, T., 14, 430
Bayes, teorema de, 61
Bernoulli, distribución de, 392
—, experimento de, 392
BERNOULLI, J., 14, 529
—, tiradas de, 393
BERNSTEIN, S. N., 68
$\beta(p,q)$ (*distribución beta*), 455
β (*σ-álgebra de Borel*), 103
β^2 (*σ-álgebra de Borel del plano*), 150
β^n (*σ-álgebra de Borel en $\mathbb{R}^n$*), 154
Beta, distribución, 455
Binomial, distribución, 393
—, tablas de la distribución, 549
— negativa, distribución, 407
Bondad de ajuste para la curva general de regresión, 336
— — — — — recta de regresión, 321
BOREL, E., 7, 10, 16, 533
BOX, G., 8
BRAVAIS, A., 493

c.p. (*convergencia en probabilidad*), 518
c.s. (*convergencia casi segura*), 517
Cálculo de probabilidades, orígenes, 13
Cambio de origen y escala de una variable aleatoria, 251
CARDANO, G., 14
Casi segura, convergencia, 517
Cauchy, distribución de, 461
CHEBYSHEV, P. L., 526, 529
Chebyshev, desigualdad de, 278
—, teorema de, 526
Cobb-Douglas, distribución de, 445
Coeficiente de apuntamiento de Fisher, 274
— de asimetría de Fisher, 273
— de correlación lineal entre dos variables aleatorias, 298
— — — muestral, 531
— de variación de una variable aleatoria, 264
Combinaciones, 49
— con repetición, 51
Combinatoria, nociones de, 47
Composición de experimentos, 70
CONDORCET, Marqués de, 15
Convergencia casi segura, 517
— de los momentos muestrales, 530
— en distribución, 523
— en ley, 522
— en media, 522
— en media cuadrática, 522
— — — de orden r, 521
— en probabilidad, 518

CORNING, H., 13
Corrección por continuidad de Fisher, 540
Covarianza entre dos variables aleatorias, 293
— muestral, 531
Criterio de los mínimos cuadrados, 314
Cuantil de una variable aleatoria, 268
— de una variable aleatoria continua, 269
Cuartiles de una variable aleatoria, 270
Curva general de regresión, 315, 334
— logística, 464

DARWIN, CH. R., 15, 313
DE GROOT, M. H., 3
DE MOIVRE, A., 14, 437, 535
De Moivre-Laplace, teorema de, 535
Deciles de una variable aleatoria, 270
Δ (*diferencia simétrica de sucesos*), 26
Desigualdad de Chebyshev, 278
Desviación absoluta media respecto a la mediana, 271
— típica de una variable aleatoria, 262
Determinista, experimento, 5, 21
—, fenómeno, 21
Diagrama de árbol en experimentos compuestos, 87
Diferencia de sucesos, 25
— simétrica de sucesos, 26
Distribución asintótica, 523
— beta, 455
— binomial, 394
— —, tablas de, 549
— — negativa, 407
— bivariante de probabilidad, 164
—, convergencia en, 523
— de Bernoulli, 392
— de Cauchy, 461
— — — estándar, 461
— de Cobb-Douglas, 445
— de Laplace-Gauss, 436
— de Pareto, 459
— de Poisson, 398
— — —, tablas de, 554
— de probabilidad, parámetros de una, 387
— — — univariante, 110
— degenerada, 390
— exponencial negativa, 452
— — —, falta de memoria, 454
— F de Snedecor, 482
— — — — descentrada, 486
— — — —, tablas de, 562
— gamma, 448
— geométrica, 403
— —, falta de memoria de la, 405
— hipergeométrica, 411
— — multivariante, 419
— ji-cuadrado, 474
— —, tablas de, 560
— — con n grados de libertad, 474
— — descentrada, 478
— leptocúrtica, 274
— logarítmico-normal, 445
— logística, 464
— mesocúrtica, 274
— multinomial, 416
— normal, 436
— —, tablas de, 559
— — bidimensional, 503
— — — estándar, 503
— — estándar, 439
— — n-dimensional, 490
— — — estándar, 489
— n-variante de probabilidad, 165
— platicúrtica, 274
— t de Student, 478
— — — — descentrada, 481
— — — —, tablas de, 561
— uniforme continua, 429
— — discreta, 388
Distribuciones tipo, 387
DURKHEIM, E., 19

Econometría, 16
EINSTEIN, A., 4

$\varepsilon(a)$ (*distribución exponencial negativa*), 452
Espacio muestral, 22
— — continuo, probabilidad en un, 43
— — finito, probabilidad en un, 38
— — — y equiprobable, probabilidad en un, 39
— — infinito numerable, probabilidad en un, 41
— — producto, 70
— — —, sucesos equivalentes en un, 71
— probabilístico, 33
— probabilizable, 31
Esperanza condicionada, definición, 302
— —, propiedades, 307
— de la suma de variables aleatorias, 290
— de una función de una variable aleatoria bidimensional, 288
— — — — de una variable aleatoria continua, 247
— — — — de una variable aleatoria discreta, 247
— — — — de una variable aleatoria n-dimensional, 289
— de una matriz aleatoria, definición, 343
— — — — —, propiedades, 343
— de una variable aleatoria compleja, 360
— — — — — continua, 245
— — — — — discreta, 244
— — — — — n-dimensional, 345
— — — — —, propiedades, 250
— del producto de variables aleatorias independientes, 291
Estadística bayesiana, 61
—, definición, 2, 6
— descriptiva, 6
— teórica, 6
— y economía, 19
η (*razón de correlación*), 335
EUCLIDES, 9
Experimento aleatorio, 5, 21, 22
— compuesto, 70
— —, diagrama de árbol en un, 87
— —, equiprobabilidad en un, 84
— —, probabilidad en un, 73
— —, probabilidades de transición en un, 78
— —, probabilidades marginales en un, 75, 80
— de Bernoulli, 392
— determinista, 5, 21
Experimentos físicamente independientes, 69
Exponencial negativa, distribución, 452

F de Snedecor descentrada, distribución, 486
— — —, distribución, 482
— — —, tablas de la distribución, 562
F_{n_1,n_2} (*distribución F de Snedecor*), 482
Falta de memoria de la distribución exponencial negativa, 454
— — — de la distribución geométrica, 405
FELLER, W., 40
Fenómeno aleatorio, 21
— determinista, 21
FERMAT, P. DE, 14
Φ (*función de distribución normal estándar*), 440
FINETTI, B. DE, 7
Fisher, corrección por continuidad, 540
FISHER, R. A., 15, 17
Frecuencia absoluta de un suceso, 31
— relativa condicionada de un suceso, 57
— — de un suceso, 31
Función característica de la suma de variables independientes, 364

— — de una variable aleatoria n-dimensional, 373
— — de una variable aleatoria unidimensional, 360
— —, teorema de inversión, 370
— —, teorema de unicidad, 370
— — y momentos de una variable aleatoria n-dimensional, 380
— — — — de una variable unidimensional, 367
Función de densidad, caracterización, 127
— — — condicionada por un suceso elemental, 193
— — — de una variable aleatoria n-dimensional continua, 184
— — — de una variable aleatoria bidimensional continua, 174
— — — de una variable aleatoria continua, 122
— — — marginal, 179, 185
— — —, propiedades, 123
Función de distribución, caracterización, 110
— — — condicionada por un suceso compuesto para variables continuas, 194
— — — — por un suceso elemental para variables continuas, 192
— — — de una variable aleatoria, 105
— — — de una variable aleatoria bidimensional, 155
— — — de una variable aleatoria mixta, 130
— — — de una variable aleatoria n-dimensional, 165
— — — marginal, 159, 166
— — —, propiedades, 106
Función de probabilidad, caracterización, 113
— — — condicionada por un suceso compuesto, 190
— — — — por un suceso elemental, 187
— — — de una variable aleatoria n-dimensional discreta, 173
— — — de una variable aleatoria bidimensional discreta, 166
— — — de una variable aleatoria discreta, 113
— — — marginal, 170
— de probabilidad/densidad de una variable mixta, 130
— generatriz de momentos de una variable aleatoria n-dimensional, 376
— — — — de una variable aleatoria unidimensional, 362

G(p) (*distribución geométrica*), 403
GALILEO, 14
GALTON, F., 15, 313, 445, 493
γ_1 (*coeficiente de asimetría de Fisher*), 273
γ_2 (*coeficiente de apuntamiento de Fisher*), 274
$\gamma(a,p)$ (*distribución gamma*), 448
$\Gamma(p)$ (*función euleriana*), 436
Gamma, distribución, 448
GAUSS, K. F., 14, 17, 437
Geométrica, distribución, 403
GNEDENKO, G. V., 526
GOSSET, W. S. (*Student*), 400, 409, 479
Grados de libertad de la distribución ji-cuadrado, 474
— — — de una t de Student, 478
Grandes números, ley débil de los, 526
— —, ley fuerte de los, 533
GRAUNT, J., 13, 17
GRAYBILL, F. A., 3

$\mathcal{H}(N,n,p)$ (*distribución hipergeométrica*), 411
$\mathcal{H}(N,m,p_1,\ldots,p_n)$ (*distribución hipergeométrica multivariante*), 419

HALLEY, E., 13, 17
Hipergeométrica, distribución, 411
— multivariante, distribución, 419
HUYGENS, CH., 14

Igualdad de sucesos, 27
$\subset$ (*inclusión de sucesos*), 27
Inclusión de sucesos, 27
Incorrelación entre dos variables aleatorias, 299
Independencia de dos variables aleatorias, 199
— — — — —, caracterización de la, 204
— — — — — continuas, 201
— — — — — discretas, 200
— — — — — y distribuciones condicionadas, 210
— de n variables aleatorias, 213
— — — — —, caracterización de la, 215
— — — — — continuas, 214
— — — — — discretas, 214
— dos a dos de sucesos, 67
— estadística de sucesos, 63
— — y probabilidad condicionada, 63
— física de experimentos, 69
— mutua de sucesos, 67
$\cap$ (*intersección de sucesos*), 25
Intersección de sucesos, 25

JEFFREYS, H., 7
χ^2 (*distribución ji-cuadrado*), 474
Ji-cuadrado con n grados de libertad, distribución, 474
— descentrada, distribución, 478
—, distribución, 474
—, tablas de la distribución, 560

KALBFLEISCH, J. G., 40
KENDALL, M. G., 3, 11, 13, 15
KEYNES, J. M., 7
Khinchin, teorema de, 528
KOLMOGOROV, A., 7, 10, 16, 32, 526, 533
Kolmogorov, teorema de, 534
KRUSKAL, W. H., 3

$\mathcal{L}$ (*convergencia en ley*), 522
LANG, A., 19
LAPLACE, P. S. DE, 4, 13, 14, 17, 34, 430, 437, 535
Laplace-Gauss, distribución de, 436
LEBESGUE, H., 16
LÈVY, P., 536
Ley, convergencia en, 522
— de la regularidad estadística, 22, 32
— débil de los grandes números, 526
— fuerte de los grandes números, 533
Límite central, teorema del, 535
LINDEBERG, J. W., 536
LOBACHEVSKI, N. I., 9
Logarítmico-normal, distribución, 445
Logística, curva, 464
—, distribución, 464
$\log N(\mu, \sigma)$ (*distribución logarítmico-normal*), 445
LYAPUNOV, A. M., 535

$\mathcal{M}(m, p_1, \ldots, p_n)$ (*distribución multinomial*), 416
MARKOV, A. A., 526
Markov, teorema de, 277
MARTÍN-GUZMÁN, M. P., 3
MARTÍN PLIEGO, F. J., 3
MARTÍNEZ SALAS, J., 40
Matriz de varianzas-covarianzas de una variable aleatoria n-dimensional, definición, 345
m.c. (*convergencia en media cuadrática*), 522
Media, convergencia en, 521
— cuadrática, convergencia en, 522
— de orden r, convergencia en, 521
Mediana de una variable aleatoria, 270
Medidas de forma de una variable aleatoria, 272

MENDEL, G., 17
Método axiomático, 9
— deductivo, 11
— estadístico, 7
Mínimos cuadrados, criterio de los, 314
MISES, R. VON, 7
Moda de una variable aleatoria, 271
Momentos absolutos de una variable aleatoria, 257
— centrados respecto de un punto de una variable aleatoria, 256
— centrales de una variable aleatoria, 256
— — de una variable aleatoria bidimensional, 292
— condicionados centrales, 310
— — no centrales, 309
— factoriales de una variable aleatoria, 257
— muestrales centrales, 531
— —, convergencia de los, 530
— — no centrales, 530
— no centrales de una variable aleatoria, 256
— — — de una variable aleatoria bidimensional, 292
MONTMORT, 409
MOOD, A., 3
m.r. (*convergencia en media de orden r*), 521
Multinomial, distribución, 416

$N(\mu, \sigma)$ (*distribución normal*), 436
$N_n(0,1)$ (*distribución normal n-dimensional estándar*), 489
$N_n(\mu, \Sigma)$ (*distribución normal n-dimensional*), 490
Normal bidimensional, distribución, 503
— — estándar, distribución, 503
—, distribución, 436
— estándar, distribución, 439
— —, tablas de la distribución, 549
— n-dimensional, distribución, 490
— — estándar, distribución, 489
Nube de puntos asociada a una variable aleatoria bidimensional, 300
Números aleatorios, tabla de, 565

Ω (*espacio muestral*), 22
ω (*punto muestral*), 22
Operaciones entre sucesos, 23

$\mathcal{P}(\lambda)$ (*distribución de Poisson*), 398
PACCIOLI, L., 14
Parábola de regresión, 315
Parámetros de una distribución de probabilidad, 388
Pareto, distribución de, 459
Partición, 30
PASCAL, B., 14, 409
PEARSON, K., 15, 17, 313
PEÑA, D., 11
Percentiles de una variable aleatoria, 270
Permutaciones, 48
— con repetición, 50
PETTY, W., 13, 17
POINCARÉ, H., 13
POISSON, S. D., 15
Poisson, distribución de, 398
—, tablas de la distribución de, 554
Probabilidad condicionada, definición, 54
— — e independencia estadística, 63
— —, propiedades, 56
—, convergencia en, 518
—, definición de, 32
— inducida por una variable aleatoria, 104
— — por una variable aleatoria bidimensional, 153
— — por una variable aleatoria n-dimensional, 154
—, interpretación frecuentalista, 6
—, — laplaciana, 6
—, — lógica, 7
—, — subjetivista, 7
—, propiedades de la, 34

— total, teorema de la, 59
Producto de sucesos, 25
Punto muestral, 22

QUETELET, 15

Razón de correlación, 335
Recorrido de una variable aleatoria, 272
— intercuartílico, 272
Recta de regresión, 315, 318
Regla de la multiplicación, 58
— de Laplace, 39
Regresión, antecedentes históricos, 313
—, curva general de, 315, 334
—, parábola de, 315
—, recta de, 315, 318
Regularidad estadística, ley de la, 22, 32
Relaciones entre sucesos, 27
RÍOS, S., 9, 10
ρ (*coeficiente de correlación lineal*), 298

SAVAGE, L. I., 7
σ-álgebra, átomos de una, 30
— de Boole, definición, 28
— — —, propiedades, 29
— de Borel, 103
— — — del plano, 150
— — — en $\mathbb{R}^n$, 154
Simulación de una variable aleatoria continua, 434
— de una variable aleatoria discreta, 435
Sucesión de variables aleatorias, 516
Suceso, 23
— complementario, 23
— compuesto, 23
— contrario, 24
— elemental, 23
—, frecuencia absoluta de un, 31
—, — relativa de un, 31
— imposible, 23
—, ocurrencia de un, 23
— seguro, 23
Sucesos, diferencia de, 25
—, diferencia simétrica de, 26
— disjuntos, 27
— estocásticos, 31
—, igualdad de, 27
—, inclusión de, 27
— incompatibles, 27
—, intersección de, 25
— mutuamente excluyentes, 27
—, operaciones entre, 23
—, producto de, 25
—, relaciones entre, 27
—, unión de, 24

t de Student descentrada, distribución, 481
— — —, distribución, 478
— — —, tablas de la distribución, 561
t_n (*distribución t de Student*), 478
Tabla de doble entrada para una variable aleatoria bidimensional discreta, 171
Tabla de la distribución binomial, 549
— — — — de Poisson, 554
— — — — F de Snedecor, 562
— — — — ji-cuadrado, 560
— — — — normal estándar, 559
— — — — t de Student, 561
— de números aleatorios, 565
TARTAGLIA, N., 14
Teorema de Bayes, 61
— de Chebyshev, 526
— de De Moivre-Laplace, 535
— de inversión para funciones características, 370, 380
— de Khinchin, 528
— de Kolmogorov, 534
— de la probabilidad total, 59
— de Markov, 277
— de unicidad para funciones características, 370
— del límite central, 535
Tipificación de una variable aleatoria, 267

Tiradas de Bernoulli, 393
Transformación de una variable aleatoria, 136
— de una variable aleatoria continua, 139
— de una variable aleatoria discreta, 138
Transformaciones de una variable aleatoria n-dimensional, 216
— de variables bidimensionales continuas, 219
— de variables bidimensionales discretas, 217
— de variables n-dimensionales continuas, 221
— de variables n-dimensionales discretas, 228
Truncamiento de una variable aleatoria, 132

$U(a,b)$ (*distribución uniforme continua*), 429
Uniforme continua, distribución, 429
— discreta, distribución, 388
$\cup$ (*unión de sucesos*), 24
Unión de sucesos, 24
— — —, probabilidad de la, 35

$\emptyset$ (*suceso imposible*), 23
Variable aleatoria bidimensional continua, 174
— — — —, función de densidad, 174
— — —, definición, 149
— — — discreta, 166
— — — —, función de probabilidad, 167
— — — —, tabla de doble entrada, 171
— — —, esperanza de una función de una, 288
— — —, función de distribución de una, 155
— — —, momentos centrales, 292
— — —, — no centrales, 292
— — —, probabilidad inducida por una, 153
— —, cambio de origen y escala, 251
— — compleja, 358
— — —, esperanza de, 360
— — continua, cambio de variable, 139
— — —, definición, 121
— — —, esperanza de, 245
— — —, función de densidad, 122
— — —, simulación de una, 434
— —, cuantil de una, 268
— —, definición, 94
— —, definición equivalente de, 103
— — discreta, cambio de variable, 138
— — —, definición, 113
— — —, esperanza de, 244
— — —, función de pobabilidad de una, 113
— — —, simulación de una, 435
— —, función de distribución de una, 105
— —, mediana de una, 270
— —, medidas de forma de una, 272
— — mixta, definición, 128
— — —, función de distribución de una, 130
— — —, función de probabilidad/densidad, 130
— —, momentos absolutos de una, 257
— —, — centrales de una, 256
— —, — factoriales de una, 257
— —, — no centrales de una, 256
— — n-dimensional, cambio de variable, 216
— — — continua, 184
— — — —, función de densidad de una, 184
— — —, definición, 154
— — — discreta, 173
— — — —, función de probabilidad, 173
— — —, esperanza de una, 345
— — —, — de una función de una, 289

. función característica, 373
. generatriz de momentos, 376
. matriz de varianzas-covarianzas, 345
. probabilidad inducida por una, 154
. percentiles de una, 270
. probabilidad inducida por una, 104
. tipificación de, 267
. transformación de una, 135
. truncamiento de una, 132
unidimensional, función característica, 360
. generatriz de momentos, 362
. varianza de una, 259, 261
Variables aleatorias bidimensionales continuas, transformaciones de, 219
discretas, transformaciones de, 217
idénticamente distribuidas, 112
n-dimensionales continuas, transformaciones de, 221
discretas, transformaciones de, 228
Variaciones, 47
con repetición, 49
Varianza condicionada, 311
de la suma de variables aleatorias, 296
de una variable aleatoria, definición, 259, 261
. propiedades, 262
residual para la curva general de regresión, 336
para la recta de regresión, 320
Vector aleatorio bidimensional, 150
n-dimensional, 154
VERHULST, P. F., 464

ÍNDICE

Prefacio VII

1. **Introducción** 1
   - 1.1. Una aproximación al concepto de Estadística 2
   - 1.2. El método estadístico 7
   - 1.3. La Estadística como rama de las matemáticas 8
   - 1.4. Desarrollo histórico del Cálculo de Probabilidades y de la Estadística 11
   - 1.5. Estadística y Ciencias Sociales 16

2. **Concepto de probabilidad** 21
   - 2.1. Fenómenos aleatorios 21
   - 2.2. Operaciones y relaciones entre sucesos. Concepto de σ-álgebra . 23
     - 2.2.1. Operaciones entre sucesos 23
     - 2.2.2. Relaciones entre sucesos 27
     - 2.2.3. Concepto de σ-álgebra 28
   - 2.3. Definición de probabilidad 31
   - 2.4. Propiedades de las probabilidades 34
   - 2.5. Algunos tipos de espacios probabilísticos 37
     - 2.5.1. Probabilidad en un espacio muestral discreto 37
     - 2.5.2. Probabilidad en un espacio muestral continuo 43
   - 2.6. Ejercicios propuestos 44

3. **Probabilidad condicionada e independencia estadística** 53
   - 3.1. Probabilidad condicionada: definición y propiedades 53
   - 3.2. Teorema de la probabilidad total y teorema de Bayes 59
   - 3.3. Independencia de sucesos 62
   - 3.4. Composición de experimentos. Probabilidad en el espacio producto 69
     - 3.4.1. Asignación de probabilidades en experimentos compuestos 73
     - 3.4.2. Equiprobabilidad en experimentos compuestos 84
   - 3.5. Ejercicios propuestos 88

4. **Variables aleatorias unidimensionales** 93
4.1. Definición de variable aleatoria 93
4.2. σ-álgebra de Borel. Probabilidad inducida por una variable aleatoria 103
4.3. Función de distribución de una variable aleatoria. Propiedades . 105
4.4. Variables aleatorias discretas 113
4.5. Variables aleatorias continuas 121
4.6. Variables aleatorias mixtas 128
4.7. Truncamiento de variables aleatorias 132
4.8. Transformaciones de una variable aleatoria 135
4.8.1. Cambio de variable para variables discretas 138
4.8.2. Cambio de variable para variables continuas 139
4.9. Ejercicios propuestos 143

5. **Variables aleatorias n-dimensionales** 149
5.1. Definición de variable aleatoria n-dimensional 149
5.2. Función de distribución. Propiedades 155
5.3. Variables n-dimensionales discretas 166
5.4. Variables n-dimensionales continuas 174
5.5. Distribuciones condicionadas 185
5.5.1. Distribuciones discretas condicionadas 187
5.5.2. Distribuciones continuas condicionadas 191
5.6. Independencia de variables aleatorias 198
5.7. Transformaciones de una variable n-dimensional 216
5.7.1. Cambio de variable para variables bidimensionales discretas 217
5.7.2. Cambio de variable para variables bidimensionales continuas 219
5.8. Ejercicios propuestos 230

6. **Características de una variable aleatoria unidimensional** 241
6.1. Esperanza matemática 243
6.1.1. Propiedades de la esperanza 250
6.2. Momentos de una variable aleatoria 256
6.2.1. Momentos centrales y no centrales 258
6.2.2. Varianza y coeficiente de variación 261
6.2.3. Tipificación de una variable 267
6.3. Otras medidas de posición, dispersión y forma 268
6.4. Teorema de Markov y desigualdad de Chebyshev 277
6.5. Ejercicios propuestos 280

7. **Características de una variable aleatoria n-dimensional** 287
7.1. Momentos de una variable aleatoria n-dimensional 287
7.2. Esperanza condicionada 302
7.3. Momentos condicionados 309
7.4. Regresión y correlación 313

7.4.1. Regresión lineal 316
7.4.2. Regresión general 333
7.5. Esperanza y matriz de varianzas-covarianzas de una variable aleatoria n-dimensional 342
7.6. Ejercicios propuestos 349

8. **Funciones características** 357
8.1. Variable aleatoria compleja. Definición y esperanza matemática . 358
8.2. Función característica de una variable aleatoria unidimensional. Propiedades 360
8.3. Función característica de una variable aleatoria n-dimensional. Propiedades 373
8.4. Ejercicios propuestos 383

9. **Modelos de distribuciones discretas** 387
9.1. Distribución uniforme discreta 388
9.2. Distribución de Bernoulli 392
9.3. Distribución binomial 393
9.4. Distribución de Poisson 398
9.5. Distribución geométrica 403
9.6. Distribución binomial negativa 407
9.7. Distribución hipergeométrica 411
9.8. Distribuciones discretas n-variantes 415
9.8.1. Distribución multinomial 416
9.8.2. Distribución hipergeométrica multivariante 419
9.9. Ejercicios propuestos 423

10. **Modelos de distribuciones continuas univariantes** 429
10.1. Distribución uniforme continua 429
10.2. Distribución normal 436
10.3. Distribución logarítmico-normal 445
10.4. Distribución gamma 448
10.5. Distribución exponencial negativa 452
10.6. Distribución beta 455
10.7. Distribución de Pareto 459
10.8. Distribución de Cauchy 461
10.9. Distribución logística 464
10.10. Ejercicios propuestos 467

11. **Distribuciones relacionadas con la normal** 473
11.1. Distribución χ^2 de Pearson 473
11.2. Distribución t de Student 478
11.3. Distribución F de Snedecor 482
11.4. Ejercicios propuestos 486

12. **Distribución normal n-dimensional** 489
12.1. Distribución normal n-dimensional 489
12.2. Distribución normal bidimensional 502
12.3. Ejercicios propuestos 507

13. **Leyes de los grandes números y Teorema del límite central** . . 513
13.1. Convergencia de sucesiones de variables aleatorias 516
13.1.1. Convergencia casi segura 517
13.1.2. Convergencia en probabilidad 518
13.1.3. Convergencia en media de orden r 521
13.1.4. Convergencia en ley o en distribución 522
13.2. Leyes de los grandes números 526
13.2.1. Leyes débiles de los grandes números 526
13.2.2. Convergencia en probabilidad de los momentos muestrales 530
13.2.3. Leyes fuertes de los grandes números 533
13.3. Teorema del límite central 535
13.3.1. Consecuencias del teorema del límite central 538
13.3.2. Corrección por continuidad de Fisher 540
13.4. Ejercicios propuestos 544

Anexo. Tablas Estadísticas 547

Glosario 567

Impreso en el mes de junio de 1994
en Talleres Gráficos DUPLEX, S. A.
Ciudad de Asunción, 26
08030 Barcelona

Este libro es el primero de dos que se dedicarán al estudio del Cálculo de Probabilidades y de la Inferencia Estadística.

Su contenido incluye los desarrollos y aplicaciones más usuales del Cálculo de Probabilidades que son de interés para las Ciencias aplicadas. Se aborda, de este modo, la construcción axiomática de la teoría de la probabilidad, dedicando una especial atención al manejo de las herramientas de cálculo habituales (probabilidad condicionada, teorema de la probabilidad total, teorema de Bayes, composición de experimentos, etc.).

El resto del libro estudia las variables aleatorias, deteniéndose en sus principales características; se suministra un repertorio cuidadosamente seleccionado de las distribuciones de probabilidad más usuales, para las que se analizan los resultados más importantes. Por último, se dedican dos capítulos a las funciones características y al comportamiento asintótico de las distribuciones (leyes de los grandes números y teorema del límite central), que serán muy utilizados en la Inferencia Estadística.

El libro está planteado con una evidente intención pedagógica y, en este sentido, se acompañan los resultados de una abundante colección de ejemplos y ejercicios resueltos. Cada capítulo finaliza con una amplia propuesta de ejercicios que ayudarán al lector a valorar el grado de comprensión de los temas estudiados.

941189-0

00

9 788434 420946

Ariel Economía

Ariel Economía

Martina Menguzzato Boulard y Juan José Renau Piqueras
La Dirección Estratégica de la empresa

Águeda Esteban Talaya y Eduardo Pérez Gorostegui
Prácticas de marketing

Francisco J. Maqueda Lafuente
Creación y dirección de empresas

Manuel Vela, Vicente Montesinos y Vicente Serra
Manual de contabilidad

E. Albi, C. Contreras, J. M. González-Páramo e I. Zubiri.
Teoría de la Hacienda Pública

Vicente Meneu, Eliseo Navarro y M.ª Teresa Barreira
Análisis y gestión del riesgo de interés

Alfonso G. Barbancho
Estadística teórica básica

Antonio Aznar y Francisco Javier Trívez
Métodos de predicción en economía I

Antonio Aznar y Francisco Javier Trívez
Métodos de predicción en economía II

Guillermo J. Sierra, José Moreno y Pedro Rodríguez
Cuentas anuales

Anna M. Gil Lafuente
Fundamentos de análisis financiero

Joseph E. Stiglitz
Economía

Emilio Albi Ibáñez y José Luis García Ariznavarreta
Sistema fiscal español