

Curso intermedio de PROBABILIDAD

Luis Rincón
Departamento de Matemáticas
Facultad de Ciencias UNAM
Circuito Exterior de CU
04510 México DF

Versión: Octubre 2007

Una versión actualizada del presente texto se encuentra disponible en formato electrónico en la dirección <http://www.matematicas.unam.mx/lars>

Contenido

1. Espacios de probabilidad	1
1.1. Espacios de probabilidad	1
1.2. σ -álgebras	3
1.3. Medidas de probabilidad	20
1.4. Independencia de eventos	33
1.5. Lema de Borel-Cantelli	37
1.6. Ejercicios	42
2. Variables aleatorias	57
2.1. Variables aleatorias	57
2.2. Función de distribución	68
2.3. Tipos de variables aleatorias	74
2.4. Integral de Riemann-Stieltjes	81
2.5. Características numéricas	84
2.6. Distribuciones discretas	94
2.7. Distribuciones continuas	100
2.8. Ejercicios	108
3. Vectores aleatorios	141
3.1. Vectores aleatorios	141
3.2. Distribución conjunta	144
3.3. Densidad conjunta	149
3.4. Distribución marginal	156
3.5. Distribución condicional	160

3.6.	Independencia	163
3.7.	Esperanza de una función de un vector aleatorio	168
3.8.	Covarianza	171
3.9.	Coefficiente de correlación	173
3.10.	Esperanza y varianza de un vector aleatorio	179
3.11.	Distribuciones multivariadas discretas	181
3.12.	Distribuciones multivariadas continuas	183
3.13.	Ejercicios	186
4.	Esperanza condicional	213
4.1.	Esperanza condicional	214
4.2.	Esperanza condicional: caso discreto	216
4.3.	Algunas propiedades	221
4.4.	Varianza condicional	223
4.5.	Ejercicios	226
5.	Transformaciones	229
5.1.	Transformación de una variable aleatoria	229
5.2.	Transformación de un vector aleatorio	235
5.3.	Ejercicios	252
6.	Dist. muestrales y estadísticas de orden	261
6.1.	Distribuciones muestrales	263
6.2.	Estadísticas de orden	271
6.3.	Ejercicios	281
7.	Convergencia	287
7.1.	Tipos de convergencia	287
7.2.	Relaciones entre los tipos de convergencia	297
7.3.	Dos resultados importantes de convergencia	303
7.4.	Ejercicios	306
8.	Funciones generadoras	311
8.1.	Función generadora de probabilidad	311
8.2.	Función generadora de momentos	316

8.3. Función característica	323
8.4. Ejercicios	336
9. Dos teoremas límite	347
9.1. Algunas desigualdades	347
9.2. Ley de los grandes números	352
9.3. Teorema central del límite	357
9.4. Ejercicios	360
A. Distribuciones de probabilidad	365
B. Conceptos y resultados varios	373

Prólogo

El presente texto está dirigido a estudiantes de mitad de carrera de las licenciaturas de matemáticas, actuaría, y áreas afines. Contiene el material básico para un segundo curso de probabilidad, y tiene como origen las notas de clase del curso semestral de Probabilidad II, que he impartido durante los últimos años en la Facultad de Ciencias de la UNAM.

El énfasis de este segundo curso se centra en la formalización de algunos conceptos estudiados en un primer curso de probabilidad, y en el estudio de vectores aleatorios y sus varios conceptos relacionados. El lector puede comprobar que se hace poco énfasis en las aplicaciones, y que la exposición cubre principalmente el desarrollo matemático. El objetivo es que después de este curso, el estudiante pueda continuar con facilidad con un curso de estadística matemática, de procesos estocásticos, o tal vez un curso avanzado de probabilidad o de teoría de la medida, teniendo como elementos básicos los conceptos teóricos aquí desarrollados. En particular se incluye un capítulo sobre esperanza condicional, cuyo uso y aplicación es cada vez más frecuente. También se incluye un capítulo sobre distribuciones muestrales y estadísticas de orden, con aplicaciones inmediatas en temas de la estadística matemática.

Al final de cada capítulo el lector encontrará una lista de ejercicios separados por temas. La mayoría de estos ejercicios son de tipo mecánico, algunos de ellos son muy sencillos de modo que el término EJERCICIOS me parece justo y adecuado. Pocos de estos ejercicios son originales, la mayor parte de

ellos son modificaciones de ejemplos o resultados clásicos que se encuentran en la larga literatura existente. La intención de contar con este material es la de crear confianza y soltura por parte del alumno en el manejo de los conceptos y notación involucrados. El número de ejercicios excede lo que normalmente puede realizarse en un semestre, y el objetivo que siempre tuve en mente estos años fue el tener un número suficiente de ellos para presentar algunos en clase, dejar otros para trabajo en casa, y asignar algunos otros para preguntas de examen, usando material ligeramente distinto cada semestre para evitar repeticiones. Durante la exposición de los temas el lector encontrará también algunos otros ejercicios propuestos y algunos ejemplos resueltos.

La presentación del material mantiene la estructura de las notas de clase, y creo que será particularmente útil al estudiante con poco tiempo para leer párrafos completos, y quien sólo busca una definición, un resultado, un ejemplo, un ejercicio, o tal vez orientación breve acerca de un concepto. En este sentido, el libro contiene tablas a manera de resumen, y los enunciados están enmarcados para su fácil localización. También he intentado que la notación fuera lo más simple y mínima posible. Personalmente me gustan los libros con imágenes y diagramas, y he buscado plasmar ese gusto en este texto. Este material fue escrito en \LaTeX , y las gráficas fueron elaboradas usando el paquete `PSTRICKS`, lo cual ha sido realmente un placer. Al final del texto aparece una lista de referencias que me permito sugerir al lector consultar para profundizar y a veces precisar en determinados temas. Algunos de estos textos no han sido mencionados explícitamente pero aparecen en la lista por que en algún momento he obtenido inspiración de ellos.

Agradezco sinceramente a todas aquellas personas, alumnos y profesores, quienes a través de sus comentarios y sugerencias, han contribuido al mejoramiento de este texto. Cualquier corrección o comentario acerca de este trabajo será muy bien recibido en el correo electrónico que aparece abajo. Es mi intención mantener en el futuro, hasta donde me sea posible, una versión electrónica actualizada, corregida y gratuita del presente texto. La página web donde puede obtenerse es

<http://www.matematicas.unam.mx/lars>

Por último, me parece importante mencionar que este texto ha sido posible, en gran medida, al excelente ambiente de trabajo y de libertad académica que he tenido la fortuna de encontrar en el Departamento de Matemáticas de la Facultad de Ciencias de la UNAM. Gracias a todos por su confianza y apoyo.

Luis Rincón
Diciembre 2006
Ciudad Universitaria UNAM
lars@fciencias.unam.mx

CAPÍTULO 1

Espacios de probabilidad

La teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios. Se entiende por experimento aleatorio todo aquel experimento tal que cuando se le repite bajo las mismas condiciones iniciales, el resultado que se obtiene no siempre es el mismo. A menudo, y por muy diversas razones, es necesario aceptar que no es posible predecir el resultado de un experimento particular aún cuando se le haya efectuado con anterioridad varias veces bajo las mismas condiciones iniciales, y en consecuencia se considera aleatorio. Bajo estas circunstancias, la teoría de la probabilidad tiene el objetivo de modelar matemáticamente cualquier experimento aleatorio de interés.

1.1. Espacios de probabilidad

El modelo matemático creado durante el primer tercio del siglo XX para estudiar los experimentos aleatorios es el así llamado *espacio de probabilidad*. Este modelo consiste de una terna ordenada, denotada usualmente por (Ω, \mathcal{F}, P) , en donde Ω es un conjunto arbitrario, \mathcal{F} es una σ -álgebra de subconjuntos de Ω , y P es una medida de probabilidad definida sobre \mathcal{F} . Explicamos a continuación brevemente cada uno de estos elementos.

ESPACIO MUESTRAL. El conjunto Ω es llamado *espacio muestral* o *espacio muestra*, y tiene como objetivo agrupar a todos los posibles resultados del experimento aleatorio en cuestión. No es imprescindible darle esta interpretación al conjunto Ω , y matemáticamente se le considera entonces como un conjunto arbitrario.

σ -ÁLGEBRA. Una clase o colección no vacía \mathcal{F} de subconjuntos de Ω es una σ -álgebra si es cerrada bajo las operaciones de tomar complementos y uniones numerables. El término σ -álgebra se lee “sigma-álgebra”. A los elementos de una σ -álgebra se les llama *eventos*, *sucesos*, o *conjuntos medibles*. Debido a su uso extendido, se usa el término medible, aunque tal vez lo correcto sea decir mensurable. En particular, un evento es *simple* o *elemental* si consta de a lo más un elemento de Ω , y es *compuesto* cuando consta de dos o más elementos de Ω .

MEDIDA DE PROBABILIDAD. Una función P definida sobre una σ -álgebra \mathcal{F} y con valores en el intervalo $[0, 1]$ es una *medida de probabilidad* si $P(\Omega) = 1$ y es σ -aditiva, es decir, si cumple que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n),$$

cuando A_1, A_2, \dots son elementos de \mathcal{F} que cumplen con la condición de ser ajenos dos a dos, esto es, $A_i \cap A_j = \emptyset$ para valores de i y j distintos. El número $P(A)$ representa una forma de medir la *posibilidad* de observar la ocurrencia del evento A , al efectuar una vez el experimento aleatorio. Tenemos entonces formalmente la siguiente definición.

DEFINICIÓN. (ESPACIO DE PROBABILIDAD). Un espacio de probabilidad es una terna (Ω, \mathcal{F}, P) , en donde Ω es un conjunto arbitrario, \mathcal{F} es una σ -álgebra de subconjuntos de Ω , y P es una medida de probabilidad definida sobre \mathcal{F} .

El objetivo es asociar un espacio de probabilidad al experimento aleatorio de interés. No existen reglas establecidas para ello y además la posible asig-

nación no es única, pues dependiendo del interés del observador, se puede asociar un espacio de probabilidad u otro. En este primer capítulo se estudian con más detalle los conceptos de σ -álgebra y medida de probabilidad. Empecemos con el primero.

1.2. σ -álgebras

En esta sección se estudia el concepto de σ -álgebra y se define la mínima σ -álgebra generada por una colección arbitraria de subconjuntos del espacio muestral. Recordemos nuevamente la definición de esta estructura.

DEFINICIÓN. (σ -ÁLGEBRA, ESPACIO MEDIBLE, EVENTO). Una colección \mathcal{F} de subconjuntos de Ω es una σ -álgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$.
3. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A la pareja (Ω, \mathcal{F}) se le llama espacio medible y a los elementos de \mathcal{F} se les llama eventos o conjuntos medibles.

En palabras, una σ -álgebra es una colección de subconjuntos de Ω que no es vacía y que es cerrada bajo las operaciones de tomar complemento y efectuar uniones infinitas numerables. Estas propiedades garantizan que la colección es cerrada al efectuar las operaciones usuales entre conjuntos, es decir, al tomar las operaciones de unión, intersección, complemento, diferencia, diferencia simétrica, etc. se obtienen nuevamente elementos de la misma colección.

En probabilidad elemental el conjunto Ω denota el espacio muestral o conjunto de posibles resultados de un experimento aleatorio, y los elementos de \mathcal{F} representan eventos en el experimento aleatorio. Una σ -álgebra es entonces una estructura que nos permite agrupar ciertos subconjuntos de Ω de interés, aquellos a los cuales se desea calcular su probabilidad, y esta estructura constituye el dominio de definición de una medida de probabilidad. Cuando el espacio muestral es finito normalmente se toma como σ -álgebra el conjunto potencia de Ω , pero para espacios muestrales más generales no siempre puede tomarse esa estructura tan grande, y deben considerarse entonces σ -álgebras más pequeñas, es por ello que se estudian estas estructuras. En general existen varias σ -álgebras que pueden asociarse a un conjunto cualquiera no vacío Ω como se muestra a continuación.

EJEMPLO Sea Ω un conjunto cualquiera no vacío. Es inmediato comprobar que cada una de las siguientes colecciones es una σ -álgebra de subconjuntos de Ω . La σ -álgebra del primer inciso es la σ -álgebra más pequeña que podemos asociar a un conjunto cualquiera Ω , y la σ -álgebra del último inciso es la más grande.

- a) $\mathcal{F}_1 = \{\emptyset, \Omega\}$.
- b) $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$, en donde $A \subseteq \Omega$.
- c) $\mathcal{F}_3 = 2^\Omega$, conjunto potencia.

EJEMPLO. Sean A y B subconjuntos de Ω tales que $A \subseteq B$. La siguiente colección es una σ -álgebra de subconjuntos de Ω que contiene explícitamente a los conjuntos A y B . ¿Puede usted verificar tal afirmación con la ayuda de un diagrama de Venn?

$$\mathcal{F} = \{\emptyset, A, B, A^c, B^c, B - A, (B - A)^c, \Omega\}$$

EJERCICIO. Sea Ω un conjunto no numerable. Demuestre que la colección \mathcal{F} dada por $\{A \subseteq \Omega : A \text{ o } A^c \text{ es finito o numerable}\}$ es una σ -álgebra.

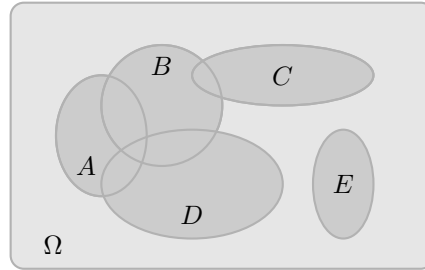


Figura 1.1: Una σ -álgebra es una colección $\mathcal{F} = \{A, B, C, D, E, \dots\}$ de subconjuntos que no es vacía y que es cerrada bajo complementos y uniones numerables.

En la Figura 1.1 puede observarse una representación gráfica de una σ -álgebra como una colección de subconjuntos de Ω . En la sección de ejercicios se pueden encontrar algunos otros ejemplos de σ -álgebras. El uso de la letra \mathcal{F} para denotar una σ -álgebra proviene del nombre en inglés “field” que significa campo. A menudo se usa también el término σ -campo en lugar de σ -álgebra. Observe con cuidado el uso y significado de los símbolos de contención y pertenencia: $A \subseteq \Omega$ y $A \in \mathcal{F}$. Demostraremos a continuación algunas otras propiedades generales de las σ -álgebras.

PROPOSICIÓN. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω . Entonces

1. $\emptyset \in \mathcal{F}$.

2. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

3. Si $A, B \in \mathcal{F}$, entonces $A - B \in \mathcal{F}$, y $A \Delta B \in \mathcal{F}$.

Demostración.

1. Como $\Omega \in \mathcal{F}$ y \mathcal{F} es una colección cerrada bajo complementos, en-

tonces $\Omega^c = \emptyset \in \mathcal{F}$.

2. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $A_1^c, A_2^c, \dots \in \mathcal{F}$. Por lo tanto $\bigcup_{n=1}^{\infty} A_n^c \in \mathcal{F}$. Tomando complementos y usando las leyes de De Morgan se obtiene el resultado.
3. Estas proposiciones se siguen de lo demostrado antes y de las definiciones $A - B := A \cap B^c$, y $A \Delta B := (A - B) \cup (B - A)$.

□

La proposición anterior establece entonces que las σ -álgebras son estructuras también cerradas bajo las operaciones de diferencia e intersecciones numerables. En la sección de ejercicios pueden encontrarse algunas otras definiciones de σ -álgebra equivalentes a la que hemos enunciado, y que involucran las operaciones de la proposición anterior. Una operación de particular importancia es aquella en la que se intersectan dos σ -álgebras produciendo una nueva σ -álgebra, este es el contenido del siguiente resultado.

PROPOSICIÓN. La intersección de dos σ -álgebras es una σ -álgebra.

Demostración. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de un mismo Ω . Entonces $\mathcal{F}_1 \cap \mathcal{F}_2$ es aquella colección de subconjuntos de Ω cuyos elementos pertenecen tanto a \mathcal{F}_1 como a \mathcal{F}_2 . Demostraremos que $\mathcal{F}_1 \cap \mathcal{F}_2$ es una σ -álgebra.

- a) Como \mathcal{F}_1 y \mathcal{F}_2 son σ -álgebras, entonces $\Omega \in \mathcal{F}_1$ y $\Omega \in \mathcal{F}_2$. Por lo tanto $\Omega \in \mathcal{F}_1 \cap \mathcal{F}_2$.
- b) Sea A un elemento en $\mathcal{F}_1 \cap \mathcal{F}_2$. Entonces $A \in \mathcal{F}_1$ y $A \in \mathcal{F}_2$. Por lo tanto $A^c \in \mathcal{F}_1$ y $A^c \in \mathcal{F}_2$, es decir, $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$.
- c) Sea A_1, A_2, \dots una sucesión de elementos en la intersección $\mathcal{F}_1 \cap \mathcal{F}_2$. Entonces $A_1, A_2, \dots \in \mathcal{F}_1$ y $A_1, A_2, \dots \in \mathcal{F}_2$. Por lo tanto $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_1$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_2$, es decir, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_1 \cap \mathcal{F}_2$. □

Hemos entonces comprobado que si \mathcal{F}_1 y \mathcal{F}_2 son dos σ -álgebras de un mismo conjunto Ω , entonces $\mathcal{F}_1 \cap \mathcal{F}_2$ es nuevamente una σ -álgebra de subconjuntos de Ω , naturalmente más pequeña que \mathcal{F}_1 y \mathcal{F}_2 en el sentido $\mathcal{F}_1 \cap \mathcal{F}_2 \subseteq \mathcal{F}_1, \mathcal{F}_2$. La siguiente pregunta consiste en verificar si la unión de dos σ -álgebras produce nuevamente una σ -álgebra. En este caso la respuesta es negativa. En general no es cierto que la unión de dos σ -álgebras produce una nueva σ -álgebra. Véanse por ejemplo los ejercicios 9 y 10 a este respecto. Por otro lado se puede extender la validez de la proposición recién demostrada a intersecciones más generales como indica el siguiente resultado.

PROPOSICIÓN. La intersección finita, infinita numerable o bien arbitraria de σ -álgebras es nuevamente una σ -álgebra.

Demostración. Sea T un conjunto arbitrario distinto del vacío. Suponga que para cada t en T se tiene una σ -álgebra \mathcal{F}_t de subconjuntos de Ω . Sea $\mathcal{F} = \bigcap_{t \in T} \mathcal{F}_t$. Siguiendo los mismos pasos que en la demostración anterior es fácil probar que \mathcal{F} es una σ -álgebra. Observe que como T es un conjunto arbitrario, la σ -álgebra \mathcal{F} es efectivamente una intersección arbitraria de σ -álgebras. \square

Lo demostrado anteriormente garantiza que la siguiente definición tiene sentido.

DEFINICIÓN. (σ -ÁLGEBRA GENERADA). Sea \mathcal{C} una colección no vacía de subconjuntos de Ω . La σ -álgebra generada por \mathcal{C} , denotada por $\sigma(\mathcal{C})$, es la colección

$$\sigma(\mathcal{C}) = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ es } \sigma\text{-álgebra y } \mathcal{C} \subseteq \mathcal{F} \}.$$

Es decir, la colección $\sigma(\mathcal{C})$ es la intersección de todas aquellas σ -álgebras que contienen a \mathcal{C} . Por la proposición anterior sabemos que $\sigma(\mathcal{C})$ es una

σ -álgebra. A $\sigma(\mathcal{C})$ también se le llama *mínima σ -álgebra generada por \mathcal{C}* , y el adjetivo *mínima* es claro a partir del hecho de que es la σ -álgebra más pequeña que contiene a la colección \mathcal{C} . Es decir, si \mathcal{F} es una σ -álgebra que contiene a \mathcal{C} , entonces forzosamente $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. Observe que $\mathcal{C} \subseteq \sigma(\mathcal{C})$ pues a la colección \mathcal{C} se le han añadido posiblemente algunos otros subconjuntos para convertirla en la σ -álgebra $\sigma(\mathcal{C})$.

EJEMPLO. Sean $A, B \subseteq \Omega$ con A y B ajenos. Defina la colección $\mathcal{C} = \{A, B\}$. En general esta colección no es una σ -álgebra pero podemos añadirle algunos subconjuntos de Ω para encontrar la σ -álgebra generada por \mathcal{C} . Resulta que la mínima σ -álgebra que contiene a la colección \mathcal{C} es la siguiente. ¿Puede usted demostrar tal afirmación?

$$\sigma(\mathcal{C}) = \{\emptyset, A, B, (A \cup B)^c, A \cup B, A^c, B^c, \Omega\}.$$

Los siguientes dos resultados son proposiciones sencillas y naturales acerca de σ -álgebras generadas. Las demostraciones son cortas pero requieren algunos momentos de reflexión en una primera lectura.

PROPOSICIÓN. Sean \mathcal{C}_1 y \mathcal{C}_2 dos colecciones de subconjuntos de Ω tales que $\mathcal{C}_1 \subseteq \mathcal{C}_2$. Entonces $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$.

Demostración. Claramente $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \sigma(\mathcal{C}_2)$. Entonces $\sigma(\mathcal{C}_2)$ es una σ -álgebra que contiene a la colección \mathcal{C}_1 . Por lo tanto $\sigma(\mathcal{C}_1) \subseteq \sigma(\mathcal{C}_2)$. \square

PROPOSICIÓN. Si \mathcal{F} es una σ -álgebra, entonces $\sigma(\mathcal{F}) = \mathcal{F}$.

Demostración. Sabemos que $\mathcal{F} \subseteq \sigma(\mathcal{F})$. Por otro lado como \mathcal{F} es una σ -álgebra que contiene a \mathcal{F} , entonces $\sigma(\mathcal{F}) \subseteq \mathcal{F}$. Esto demuestra la igualdad. \square

EJERCICIO. Demuestre que $\sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$, en donde \mathcal{C} una colección de subconjuntos de Ω . ■

EJERCICIO. Demuestre que $\sigma(\mathcal{C}_1 \cup \mathcal{C}_2) = \sigma(\sigma(\mathcal{C}_1) \cup \sigma(\mathcal{C}_2))$, en donde \mathcal{C}_1 y \mathcal{C}_2 son dos colecciones no vacías de subconjuntos de Ω . ■

Otras estructuras de subconjuntos

En esta sección se presentan los conceptos de álgebra y semi-álgebra, y su relación con σ -álgebras. No estudiaremos estas estructuras con detalle pero las mencionamos porque desempeñan un papel importante en la construcción y extensión de medidas de probabilidad.

DEFINICIÓN. (ÁLGEBRA). Una colección \mathcal{A} de subconjuntos de Ω es una álgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$.
3. Si $A_1, \dots, A_n \in \mathcal{A}$, entonces $\bigcup_{k=1}^n A_k \in \mathcal{A}$.

La diferencia entre una álgebra y una σ -álgebra estriba en que para la primera se pide que sea una colección cerrada bajo uniones finitas mientras que la segunda es una colección cerrada bajo uniones infinitas numerables. Claramente toda σ -álgebra es una álgebra.

DEFINICIÓN. (SEMIÁLGEBRA). Una colección \mathcal{S} de subconjuntos de Ω es una semiálgebra si cumple las siguientes condiciones:

1. $\Omega \in \mathcal{S}$.
2. Si $A, B \in \mathcal{S}$, entonces $A \cap B \in \mathcal{S}$.
3. Si $A, A_1 \in \mathcal{S}$ son tales que $A_1 \subseteq A$, entonces existen $A_2, \dots, A_n \in \mathcal{S}$ tales que los subconjuntos A_1, \dots, A_n son ajenos dos a dos y se cumple que

$$A = \bigcup_{k=1}^n A_k.$$

Los conceptos de σ -álgebra, álgebra y semiálgebra están relacionados como se muestra en la Figura 1.2. En la sección de ejercicios se pide demostrar las implicaciones y no implicaciones que se obtienen de este diagrama.

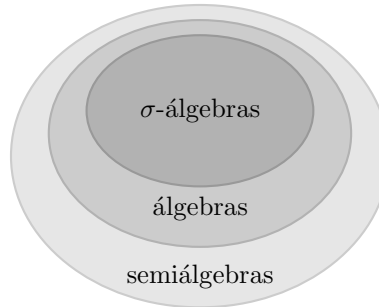


Figura 1.2: Relación general entre σ -álgebras, álgebras y semiálgebras.

A continuación se estudia un ejemplo particular de σ -álgebra de subconjuntos de números reales: la σ -álgebra de Borel.

Conjuntos de Borel

Considere la colección de todos los intervalos abiertos (a, b) de \mathbb{R} , en donde $a \leq b$. A la mínima σ -álgebra generada por esta colección se le llama σ -álgebra de Borel de \mathbb{R} , y se le denota por $\mathcal{B}(\mathbb{R})$.

DEFINICIÓN. (σ -ÁLGEBRA DE BOREL DE \mathbb{R}).

$$\mathcal{B}(\mathbb{R}) = \sigma \{ (a, b) \subseteq \mathbb{R} : a \leq b \}.$$

A los elementos de $\mathcal{B}(\mathbb{R})$ se les llama *conjuntos de Borel*, *Borelianos* o *conjuntos Borel medibles*. De esta forma se puede asociar la σ -álgebra $\mathcal{B}(\mathbb{R})$ al conjunto de números reales, y obtener así el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Se muestran a continuación algunos elementos explícitos de esta σ -álgebra.

PROPOSICIÓN. Para cualesquiera números reales $a \leq b$, los intervalos $[a, b]$, (a, ∞) , $(-\infty, b)$, $[a, b)$, $(a, b]$ y $\{a\}$, son todos elementos de $\mathcal{B}(\mathbb{R})$.

Demostración. Primeramente observe que los intervalos cerrados $[a, b]$ son conjuntos Borelianos, pues podemos escribirlos en términos de una intersección numerable de intervalos abiertos de la siguiente forma

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n} \right).$$

Observe que cada elemento de la intersección anterior es un conjunto Boreliano. Siendo $\mathcal{B}(\mathbb{R})$ una σ -álgebra, la intersección infinita es un elemento de $\mathcal{B}(\mathbb{R})$. De esta forma se concluye que cada intervalo cerrado es un conjunto

de Borel. Asimismo tenemos que

$$\begin{aligned} (a, \infty) &= \bigcup_{n=1}^{\infty} (a, a+n) \in \mathcal{B}(\mathbb{R}), \\ \text{y} \quad (-\infty, b) &= \bigcup_{n=1}^{\infty} (b-n, b) \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

Por lo tanto

$$\begin{aligned} [a, \infty) &= \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, \infty) \in \mathcal{B}(\mathbb{R}), \\ \text{y} \quad (-\infty, b] &= \bigcap_{n=1}^{\infty} (-\infty, b + \frac{1}{n}) \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

De forma análoga se puede hacer ver que los intervalos semiabiertos de la forma $[a, b)$ y $(a, b]$ son conjuntos Borelianos. Los conjuntos que constan de un solo número también son conjuntos Borelianos pues

$$\{a\} = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, a + \frac{1}{n}).$$

□

Complementos, intersecciones y uniones numerables de estos conjuntos son todos ellos Borelianos. Este hecho puede utilizarse para comprobar los siguientes resultados.

EJERCICIO. Demuestre directamente que \mathbb{N} , \mathbb{Z} y \mathbb{Q} son elementos de $\mathcal{B}(\mathbb{R})$. Demuestre además que el conjunto de números irracionales es un conjunto de Borel de \mathbb{R} . ▪

Además de la definición enunciada, existen otras formas equivalentes de generar a los conjuntos Borelianos. Este es el contenido de la siguiente proposición.

PROPOSICIÓN. Las siguientes σ -álgebras son todas idénticas a $\mathcal{B}(\mathbb{R})$.

- | | |
|------------------------------------|--|
| 1. $\sigma\{[a, b] : a \leq b\}$. | 4. $\sigma\{(a, \infty) : a \in \mathbb{R}\}$. |
| 2. $\sigma\{(a, b] : a \leq b\}$. | 5. $\sigma\{(-\infty, b) : b \in \mathbb{R}\}$. |
| 3. $\sigma\{[a, b) : a \leq b\}$. | |

Demostración. Se prueba únicamente el primer inciso, el resto de ellos se demuestra usando el mismo procedimiento. Para demostrar que $\mathcal{B}(\mathbb{R}) = \sigma\{[a, b] : a \leq b\}$ se verifican ambas contenciones. Claramente $[a, b] \in \mathcal{B}(\mathbb{R})$, por lo tanto $\{[a, b] : a \leq b\} \subseteq \mathcal{B}(\mathbb{R})$. Entonces $\sigma\{[a, b] : a \leq b\} \subseteq \mathcal{B}(\mathbb{R})$. Ahora se demuestra la contención contraria. Sabemos que $(a, b) \in \sigma\{[a, b] : a \leq b\}$ pues $(a, b) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b - \frac{1}{n}]$. Entonces $\{(a, b) : a \leq b\} \subseteq \sigma\{[a, b] : a \leq b\}$. Por lo tanto $\mathcal{B}(\mathbb{R}) \subseteq \sigma\{[a, b] : a \leq b\}$. \square

De manera equivalente se puede definir a $\mathcal{B}(\mathbb{R})$ como la mínima σ -álgebra generada por todos los subconjuntos abiertos de \mathbb{R} . En ambos casos la σ -álgebra generada es $\mathcal{B}(\mathbb{R})$.

Es natural preguntarse si la colección $\mathcal{B}(\mathbb{R})$ contiene a todos los subconjuntos de \mathbb{R} . La respuesta es negativa, es decir, puede demostrarse que existe un subconjunto de los números reales que no pertenece a la colección $\mathcal{B}(\mathbb{R})$. La construcción del tal conjunto no es sencilla, y puede obtenerse indirectamente de la siguiente forma: la colección $\mathcal{B}(\mathbb{R})$ está contenida en una clase más amplia llamada la colección de conjuntos *Lebesgue medibles* de \mathbb{R} , y se demuestra que existen subconjuntos de \mathbb{R} que no son Lebesgue medibles, y por tanto tampoco Borel medibles. Los detalles de estas afirmaciones pueden encontrarse en textos de teoría de la medida, como por ejemplo [5] o [14].

Es posible también considerar la σ -álgebra de conjuntos de Borel restringidos a una porción de los números reales como se indica a continuación.

DEFINICIÓN. Sea $A \in \mathcal{B}(\mathbb{R})$. La σ -álgebra de Borel de A , denotada por $\mathcal{B}(A)$ o por $A \cap \mathcal{B}(\mathbb{R})$, se define como sigue

$$\mathcal{B}(A) = \{A \cap B : B \in \mathcal{B}(\mathbb{R})\}.$$

No es difícil comprobar que la colección $\mathcal{B}(A)$ es efectivamente una σ -álgebra de subconjuntos de A . Observe que el nuevo conjunto total es A y no \mathbb{R} . El concepto de σ -álgebra de Borel de \mathbb{R} puede extenderse a dimensiones mayores de la siguiente forma. Considere la colección \mathcal{C} de todas los rectángulos abiertos de \mathbb{R}^2 , es decir,

$$\mathcal{C} = \{(a, b) \times (c, d) : a \leq b, c \leq d\}.$$

Se definen los conjuntos de Borel de \mathbb{R}^2 como los elementos de la mínima σ -álgebra generada por la colección \mathcal{C} , es decir, $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{C})$. De manera equivalente se puede definir $\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}))$. En forma análoga se define $\mathcal{B}(\mathbb{R}^n)$ usando productos cartesianos de intervalos.

DEFINICIÓN. (σ -ÁLGEBRA DE BOREL DE \mathbb{R}^n).

$$\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})).$$

En general el producto cartesiano de dos σ -álgebras no es una σ -álgebra de subconjuntos del espacio producto, de modo que debe anteponerse la operación σ a tal colección para convertirla en una σ -álgebra.

EJERCICIO. (σ -ÁLGEBRA PRODUCTO). Demuestre que el producto cartesiano de dos σ -álgebras no es necesariamente σ -álgebra. Esto es, suponga que $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ son dos espacios medibles. Mediante un ejemplo muestre que $\mathcal{F}_1 \times \mathcal{F}_2$ no necesariamente es una σ -álgebra de subconjuntos del espacio producto $\Omega_1 \times \Omega_2$. Se define entonces la σ -álgebra producto como

$\sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. ▪

EJERCICIO. Demuestre que la σ -álgebra $\sigma\{(a, b) \times (c, d) : a \leq b, c \leq d\}$ coincide con $\sigma(\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}))$. ▪

Sucesiones de eventos

En esta sección se estudia el concepto de convergencia de una sucesión infinita de eventos. Para enunciar tal concepto necesitaremos antes las definiciones de límite superior y límite inferior para conjuntos. Estas definiciones son análogas al caso de sucesiones numéricas como puede consultarse en un apéndice al final del texto.

DEFINICIÓN. (LÍMITE SUPERIOR E INFERIOR). Para una sucesión de eventos $\{A_n : n \in \mathbb{N}\}$, se define el límite superior y el límite inferior como sigue:

$$1. \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

$$2. \liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Tanto el límite superior como el límite inferior son operaciones bien definidas, es decir, el resultado siempre existe y es único. En cada caso, el conjunto resultante es siempre un evento, es decir, un conjunto medible. Es sencillo también comprobar que

$$\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$$

Tampoco es difícil verificar que un elemento pertenece al evento límite superior si, y sólo si, pertenece a una infinidad de elementos de la sucesión. En

algunos textos de habla inglesa el evento límite superior se escribe (A_n i.o.), en donde las letras i.o. significan “infinitely often”. Por otro lado un elemento pertenece al evento límite inferior si, y sólo si, pertenece a todos los elementos de la sucesión excepto un número finito de ellos. Con estos conceptos podemos ahora establecer la definición de convergencia de una sucesión de eventos.

DEFINICIÓN. (CONVERGENCIA DE EVENTOS). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Si existe un evento A tal que

$$\liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n = A,$$

entonces se dice que la sucesión converge al evento A , y se escribe $\lim_{n \rightarrow \infty} A_n = A$.

Para calcular el posible límite de una sucesión de eventos debemos entonces calcular el límite superior y el límite inferior, y cuando el resultado de ambas operaciones coincida, entonces a tal resultado común se le llama el límite de la sucesión.

EJEMPLO. Para cada número natural n defina el conjunto $A_n = [-1/n, 0]$ si n es impar, y $A_n = [0, 1/n]$ si n es par. Entonces $\lim_{n \rightarrow \infty} A_n = \{0\}$ pues

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} [-1/n, 1/n] = \{0\}, \\ \text{y } \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \{0\} = \{0\}. \end{aligned}$$

EJERCICIO. Sea A un evento. Demuestre que la siguiente sucesión de eventos no es convergente.

$$A_n = \begin{cases} A & \text{si } n \text{ es impar,} \\ A^c & \text{si } n \text{ es par.} \end{cases}$$

Como el ejercicio anterior muestra, no todas las sucesiones de eventos convergen. Demostramos a continuación que en particular toda sucesión monótona es convergente. Más adelante presentaremos algunos otros ejemplos concretos de sucesiones de eventos, y en la sección de ejercicios se encuentran algunos más.

PROPOSICIÓN. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión monótona de eventos.

1. Si $A_1 \subseteq A_2 \subseteq \dots$, entonces $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.
2. Si $A_1 \supseteq A_2 \supseteq \dots$, entonces $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Demostración.

1. Como la sucesión es creciente, entonces (observe el valor inicial del subíndice en las operaciones de unión e intersección),

$$\bigcup_{k=n}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k,$$

y

$$\bigcap_{k=n}^{\infty} A_k = A_n.$$

Por lo tanto

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k,$$

y

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} A_n.$$

2. El procedimiento es completamente análogo al inciso anterior. En este caso como la sucesión es decreciente se tiene que

$$\bigcap_{k=n}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k,$$

y

$$\bigcup_{k=n}^{\infty} A_k = A_n.$$

Entonces

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=1}^{\infty} A_n,$$

y

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} A_k.$$

□

El siguiente resultado establece que a partir de una sucesión de eventos puede construirse otra sucesión cuyos elementos son ajenos dos a dos, y cuya unión es la unión de la sucesión original. Este procedimiento de separación será de utilidad más adelante.

PROPOSICIÓN. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Defina

$$B_1 = A_1, \quad \text{y} \quad B_n = A_n - \bigcup_{k=1}^{n-1} A_k, \quad \text{para } n \geq 2.$$

Entonces la sucesión de eventos $\{B_n : n \in \mathbb{N}\}$ satisface las siguientes propiedades:

1. $B_n \subseteq A_n$.
2. $B_n \cap B_m = \emptyset$, si $n \neq m$.
3. $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$.

Demostración.

1. Esto es evidente a partir de la definición de B_n .
2. Sin pérdida de generalidad suponga que $n < m$, entonces

$$\begin{aligned} B_n \cap B_m &= \left(A_n - \bigcup_{k=1}^{n-1} A_k \right) \cap \left(A_m - \bigcup_{k=1}^{m-1} A_k \right) \\ &= \left(A_n \cap \bigcap_{k=1}^{n-1} A_k^c \right) \cap \left(A_m \cap \bigcap_{k=1}^{m-1} A_k^c \right) \\ &\subseteq A_n \cap A_n^c \\ &= \emptyset. \end{aligned}$$

3. Consideraremos cada contención por separado. Como cada B_n está contenido en A_n , entonces el lado izquierdo es efectivamente un subconjunto del lado derecho. Por el contrario, sea x un elemento en

$\bigcup_{n=1}^{\infty} A_n$. Entonces existe un índice n tal que $x \in A_n$. Sea n_0 el primer índice tal que $x \in A_{n_0}$ y $x \notin A_j$ para $1 \leq j \leq n_0 - 1$. Entonces $x \in A_{n_0} - \bigcup_{n=1}^{n_0-1} A_n = B_{n_0}$. Por lo tanto x pertenece a $\bigcup_{n=1}^{\infty} B_n$.

□

1.3. Medidas de probabilidad

En esta sección y en lo que resta del presente capítulo se estudian algunas propiedades de las medidas de probabilidad. Empezaremos por recordar nuevamente la definición de este concepto.

DEFINICIÓN. (MEDIDA DE PROBABILIDAD). Sea (Ω, \mathcal{F}) un espacio medible. Una medida de probabilidad es una función $P : \mathcal{F} \rightarrow [0, 1]$ que satisface

1. $P(\Omega) = 1$.
2. $P(A) \geq 0$, para cualquier $A \in \mathcal{F}$.
3. Si $A_1, A_2, \dots \in \mathcal{F}$ son ajenos dos a dos, esto es, $A_n \cap A_m = \emptyset$ para $n \neq m$, entonces $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

Entonces toda función P definida sobre una σ -álgebra \mathcal{F} , con valores en el intervalo $[0, 1]$ y que cumpla los tres postulados anteriores se le llama *medida de probabilidad* o *probabilidad axiomática*. Estos axiomas fueron establecidos por A. N. Kolmogorov en 1933. En particular, la tercera propiedad se conoce con el nombre de σ -aditividad.

EJEMPLO. (PROBABILIDAD CLÁSICA). Considere un experimento aleatorio con espacio muestral un conjunto finito Ω . Asocie a este conjunto la σ -

álgebra 2^Ω , y para cualquier subconjunto A de Ω defina $P(A) = \#A/\#\Omega$. Entonces P es una medida de probabilidad, y es llamada *probabilidad clásica*. De acuerdo a esta definición, para calcular la probabilidad de un evento es necesario entonces conocer su cardinalidad. En los inicios de la teoría de la probabilidad se consideraban únicamente modelos de este tipo, los cuales eran estudiados en el contexto de los juegos de azar. De esta forma de calcular probabilidades surgen muchos y muy variados problemas de conteo, algunos de los cuales pueden no ser fáciles de resolver. Por ejemplo, si cuatro parejas se sientan al azar en una mesa circular, ¿cuál es la probabilidad de que ninguna persona se sienta junto a su pareja? \blacksquare

EJEMPLO. Considere un experimento aleatorio con espacio muestral el conjunto de números naturales \mathbb{N} . Asocie a este conjunto la σ -álgebra $2^\mathbb{N}$. Para cualquier subconjunto A de \mathbb{N} defina

$$P(A) = \sum_{n \in A} \frac{1}{2^n}.$$

Es decir, el número natural n tiene asociada la probabilidad $1/2^n$, como se muestra en la Figura 1.3. No es difícil verificar que P es efectivamente una medida de probabilidad concentrada en el conjunto de números naturales.

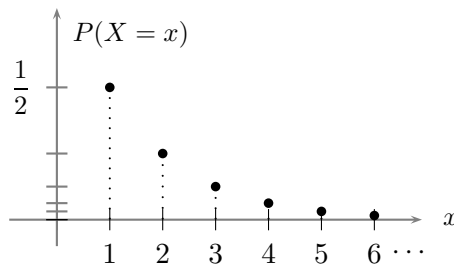


Figura 1.3: Una medida de probabilidad concentrada en los números naturales. \blacksquare

EJEMPLO. Considere el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sea $f : \mathbb{R} \rightarrow [0, \infty)$ una función no negativa y continua, tal que su integral sobre el intervalo

$(-\infty, \infty)$ es uno. La función P definida para cualquier conjunto de Borel A por la siguiente integral, es una medida de probabilidad.

$$P(A) = \int_A f(x) dx.$$

▪

EJEMPLO. (PROBABILIDAD GEOMÉTRICA). Sea $\Omega \subseteq \mathbb{R}^2$ una región tal que su área es positiva y finita. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω para los cuales el concepto de área esté bien definido. Para cada A en \mathcal{F} define $P(A) = \text{Área}(A)/\text{Área}(\Omega)$. La función P resulta ser una medida de probabilidad, y es llamada *probabilidad geométrica*. Esta definición puede extenderse a espacios de dimensión mayor de manera evidente. Un ejemplo en donde se utiliza esta forma de calcular probabilidades es el siguiente: ¿cuál es la probabilidad de que una dardo lanzado al azar sobre un tablero circular de radio unitario caiga en el círculo circunscrito de radio $1/2$? ▪

En la siguiente sección estudiaremos algunas propiedades generales que cumple toda medida de probabilidad, y a lo largo del texto consideraremos varios modelos particulares para el cálculo de probabilidades.

Propiedades elementales

A partir de los postulados enunciados en la sección anterior es posible demostrar una extensa serie de propiedades que cumplen todas las medidas de probabilidad. En esta sección se estudian algunas propiedades elementales que posiblemente ya conoce el lector, y más adelante se demuestran otras propiedades ligeramente más avanzadas.

PROPOSICIÓN. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Entonces

1. $P(\emptyset) = 0$.
2. Si $A_1, \dots, A_n \in \mathcal{F}$ son ajenos dos a dos, entonces

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k).$$

3. $P(A^c) = 1 - P(A)$.
4. Si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$.
5. Si $A \subseteq B$, entonces $P(A) \leq P(B)$.
6. $0 \leq P(A) \leq 1$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
8. $P(A \cup B) \leq P(A) + P(B)$.

Demostración.

1. Como $\emptyset = \emptyset \cup \emptyset \cup \dots$, por la σ -aditividad se tiene que $P(\emptyset) = \sum_{n=1}^{\infty} P(\emptyset)$, lo cual sucede únicamente cuando $P(\emptyset) = 0$.
2. Se toma $A_{n+1} = A_{n+2} = \dots = \emptyset$, y la igualdad se obtiene al aplicar la σ -aditividad y la propiedad anterior.
3. Se expresa a Ω como la unión disjunta $A \cup A^c$. Aplicamos P y obtenemos la igualdad requerida.
4. Escribimos $B = A \cup (B - A)$. Aplicando P obtenemos $P(B) - P(A) = P(B - A)$.
5. Como la probabilidad de cualquier evento es un número no negativo, el resultado se sigue de la propiedad anterior.

6. La primera desigualdad es el segundo axioma, y la segunda es consecuencia de la propiedad anterior cuando $B = \Omega$ y el primer axioma.
7. Descomponemos el evento $A \cup B$ como la siguiente unión de tres eventos disjuntos dos a dos: $A \cup B = (A - B) \cup (A \cap B) \cup (B - A) = (A - A \cap B) \cup (A \cap B) \cup (B - A \cap B)$. Por lo tanto $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$.
8. Esta propiedad es consecuencia de la anterior y el segundo axioma.

□

La propiedad (2) establece que las probabilidades son funciones *finitamente aditivas*, y la propiedad (5) que son funciones *monótonas*. La desigualdad (8) dice que las probabilidades son funciones *finitamente subaditivas*. Veamos algunas otras propiedades.

PROPOSICIÓN. (DESIGUALDADES DE BOOLE). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Entonces

1. $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$.
2. $P\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c)$.

Demostración.

1. Tome $B_1 = A_1$, y para $n \geq 2$ defina

$$B_n = A_n - \bigcup_{k=1}^{n-1} A_k.$$

Hemos demostrado antes que $\{B_n : n \in \mathbb{N}\}$ es una sucesión de eventos disjuntos dos a dos tales que $B_n \subseteq A_n$ y $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Por lo tanto

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &\leq \sum_{n=1}^{\infty} P(A_n). \end{aligned}$$

2. Esta desigualdad se sigue de la primera al considerar la sucesión de los complementos.

□

PROPOSICIÓN. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos.

1. Si $P(A_n) = 1$ para toda n , entonces $P(\bigcap_{n=1}^{\infty} A_n) = 1$.
2. Si $P(A_n) = 1$ para alguna n , entonces $P(\bigcup_{n=1}^{\infty} A_n) = 1$.
3. Si $P(A_n) = 0$ para alguna n , entonces $P(\bigcap_{n=1}^{\infty} A_n) = 0$.
4. Si $P(A_n) = 0$ para toda n , entonces $P(\bigcup_{n=1}^{\infty} A_n) = 0$.

Demostración.

1. Por las leyes de De Morgan y la desigualdad de Boole,

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= 1 - P\left(\bigcup_{n=1}^{\infty} A_n^c\right) \\ &\geq 1 - \sum_{n=1}^{\infty} P(A_n^c) \\ &= 1. \end{aligned}$$

2. Como $A_n \subseteq \bigcup_{n=1}^{\infty} A_n$, se tiene que $1 = P(A_n) \leq P\left(\bigcup_{n=1}^{\infty} A_n\right)$.

3. Como $\bigcap_{n=1}^{\infty} A_n \subseteq A_n$, entonces $P\left(\bigcap_{n=1}^{\infty} A_n\right) \leq P(A_n) = 0$.

4. Por la desigualdad de Boole, $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) = 0$.

□

Las propiedades (1) y (4) de la proposición anterior pueden interpretarse de la siguiente forma. Intersectar dos eventos produce en general un evento más pequeño, o por lo menos no mayor a los interseccionados. Sin embargo la propiedad (1) establece que la intersección, aún infinita, de eventos con probabilidad uno produce un evento con probabilidad uno. Análogamente, unir dos eventos produce en general un evento mayor, pero por la propiedad (4), la unión, aún infinita, de eventos con probabilidad cero tiene probabilidad cero.

Dos de las propiedades elementales más conocidas y de amplia aplicación son la fórmula de probabilidad total y la fórmula de Bayes. Estas fórmulas hacen uso de la *probabilidad condicional* de un evento A dado otro evento B definida como sigue $P(A|B) := P(A \cap B)/P(B)$, cuando $P(B) \neq 0$.

EJERCICIO. (TEOREMA DE PROBABILIDAD TOTAL). Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea $\{A_1, A_2, \dots\}$ una partición de Ω tal que cada elemento de la partición es un evento con probabilidad estrictamente positiva. Demuestre que para cualquier evento B ,

$$P(B) = \sum_{n=1}^{\infty} P(B | A_n) P(A_n).$$

▪

EJERCICIO. (TEOREMA DE BAYES). Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea A_1, A_2, \dots una partición de Ω tal que cada elemento de la partición es un evento con probabilidad estrictamente positiva. Demuestre que para cualquier evento B tal que $P(B) > 0$, y para cualquier $m \geq 1$ fijo,

$$P(A_m | B) = \frac{P(B | A_m) P(A_m)}{\sum_{n=1}^{\infty} P(B | A_n) P(A_n)}.$$

▪

EJERCICIO. (COMPLETACIÓN DE ESPACIOS). Se dice que un espacio de probabilidad (Ω, \mathcal{F}, P) es *completo* si cada vez que se tenga la situación $A \subseteq B$ con $B \in \mathcal{F}$ y $P(B) = 0$, entonces también se tiene que $A \in \mathcal{F}$ y $P(A) = 0$. Un espacio de probabilidad (Ω, \mathcal{F}, P) que no es completo puede ser completado de la siguiente forma. Se toma el mismo Ω y se define la colección $\bar{\mathcal{F}}$ de todos aquellos subconjuntos $A \subseteq \Omega$ para los cuales existan B y C en \mathcal{F} con $P(C) = 0$, tales que $B \subseteq A \subseteq B \cup C$. Para tal conjunto A se define $\bar{P}(A) = P(B)$. Entonces resulta que $(\Omega, \bar{\mathcal{F}}, \bar{P})$ es un espacio de probabilidad completo, y se llama la *completación* de (Ω, \mathcal{F}, P) . Verifique esta afirmación demostrando los siguientes incisos.

- a) $\bar{\mathcal{F}}$ es efectivamente una σ -álgebra.
- b) $\mathcal{F} \subseteq \bar{\mathcal{F}}$.

- c) La definición de $\bar{P}(A)$ no depende del subconjunto B asociado, es decir, la definición es única.
- d) \bar{P} es una medida de probabilidad sobre $\tilde{\mathcal{F}}$.
- e) $\bar{P}(A) = P(A)$, para cada A en \mathcal{F} .
- f) El espacio de probabilidad $(\Omega, \tilde{\mathcal{F}}, \bar{P})$ es completo.
- g) $(\Omega, \tilde{\mathcal{F}}, \bar{P})$ es el espacio de probabilidad completo más pequeño que contiene a (Ω, \mathcal{F}, P) , es decir, si $(\Omega, \mathcal{F}_1, P_1)$ es otro espacio de probabilidad completo tal que $\mathcal{F} \subseteq \mathcal{F}_1$ y $P_1 = P$ sobre \mathcal{F} , entonces $\tilde{\mathcal{F}} \subseteq \mathcal{F}_1$ y $\bar{P} = P_1$ sobre $\tilde{\mathcal{F}}$.

▪

Continuidad

Ahora demostraremos que las medidas de probabilidad son funciones continuas. Primero se prueba este resultado importante para dos tipos de sucesiones particulares, aquellas que son monótonas crecientes o decrecientes, y después se prueba en general. Empezaremos con el caso de sucesiones crecientes.

PROPOSICIÓN. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión no decreciente de eventos, esto es, $A_1 \subseteq A_2 \subseteq \dots$. Entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Demostración. Como $A_n \subseteq A_{n+1}$, tenemos que $P(A_n) \leq P(A_{n+1})$. Por lo tanto la sucesión numérica $\{P(A_n) : n \in \mathbb{N}\}$ es no decreciente y acotada

superiormente por uno. Entonces el límite de esta sucesión existe y el lado derecho de la igualdad tiene sentido. Defina los eventos

$$B_1 = A_1,$$

$$\text{y } B_n = A_n - A_{n-1}, \quad \text{para } n \geq 2.$$

La sucesión $\{B_n : n \in \mathbb{N}\}$ es una colección de eventos disjuntos dos a dos, y es tal que

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

Por lo tanto

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= \sum_{n=1}^{\infty} P(B_n) \\ &= P(B_1) + \sum_{n=2}^{\infty} P(B_n) \\ &= P(A_1) + \sum_{n=2}^{\infty} P(A_n - A_{n-1}) \\ &= P(A_1) + \sum_{n=2}^{\infty} P(A_n) - P(A_{n-1}) \\ &= P(A_1) + \lim_{m \rightarrow \infty} \sum_{n=2}^m P(A_n) - P(A_{n-1}) \\ &= P(A_1) + \lim_{m \rightarrow \infty} P(A_m) - P(A_1) \\ &= \lim_{m \rightarrow \infty} P(A_m). \end{aligned}$$

□

Las medidas de probabilidad también son continuas respecto de sucesiones no crecientes de eventos. Esta afirmación es el contenido del siguiente resultado que se demuestra a partir de la proposición anterior.

PROPOSICIÓN. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión no creciente de eventos, esto es, $A_1 \supseteq A_2 \supseteq \dots$. Entonces

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Demostración. Observe que si $A_n \supseteq A_{n+1}$, entonces $A_n^c \subseteq A_{n+1}^c$. Por la proposición anterior,

$$P\left(\bigcup_{n=1}^{\infty} A_n^c\right) = \lim_{n \rightarrow \infty} P(A_n^c).$$

Aplicando las leyes de De Morgan,

$$1 - P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} (1 - P(A_n)),$$

de donde se sigue inmediatamente el resultado. \square

Ahora se enuncia un resultado más fuerte. Demostraremos que las medidas de probabilidad son funciones continuas. Esta propiedad es muy útil pues permite el cálculo de probabilidades en procedimientos límite, y se encuentra siempre presente de manera implícita en toda la teoría que se desarrolla más adelante.

PROPOSICIÓN. (CONTINUIDAD DE LA PROBABILIDAD). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos convergente al evento A . Entonces

$$\lim_{n \rightarrow \infty} P(A_n) = P(A).$$

Demostración. La prueba se basa en las siguientes dos desigualdades:

- a) $\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n)$.
 b) $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n)$.

Como la sucesión de eventos es convergente al evento A , entonces el límite superior y el límite inferior son iguales a A . Se sigue entonces de las desigualdades (a) y (b) que

$$\begin{aligned} \limsup_{n \rightarrow \infty} P(A_n) &\leq P(\limsup_{n \rightarrow \infty} A_n) \\ &= P(A) \\ &= P(\liminf_{n \rightarrow \infty} A_n) \\ &\leq \liminf_{n \rightarrow \infty} P(A_n). \end{aligned}$$

De donde se concluye el resultado. Nos concentraremos ahora en demostrar las desigualdades enunciadas.

- a) Como $A_n \subseteq \bigcup_{k=n}^{\infty} A_k$, entonces

$$P(A_n) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right),$$

en donde $\{\bigcup_{k=n}^{\infty} A_k : n \in \mathbb{N}\}$ es una sucesión decreciente de eventos. Tomando el límite superior se obtiene

$$\begin{aligned}
 \limsup_{n \rightarrow \infty} P(A_n) &\leq \limsup_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\
 &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \\
 &= P\left(\lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k\right) \\
 &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\
 &= P(\limsup_{n \rightarrow \infty} A_n).
 \end{aligned}$$

b) Como $\bigcap_{k=n}^{\infty} A_k \subseteq A_n$, entonces

$$P\left(\bigcap_{k=n}^{\infty} A_k\right) \leq P(A_n),$$

en donde $\{\bigcap_{k=n}^{\infty} A_k : n \in \mathbb{N}\}$ es una sucesión creciente de eventos. Tomando el límite inferior se obtiene

$$\begin{aligned}
 \liminf_{n \rightarrow \infty} P(A_n) &\geq \liminf_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
 &= \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k\right) \\
 &= P\left(\lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k\right) \\
 &= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) \\
 &= P(\liminf_{n \rightarrow \infty} A_n).
 \end{aligned}$$

□

EJEMPLO. Se lanza un dado equilibrado una infinidad de veces. Sea A_n el evento correspondiente a obtener el evento $A = \{2, 4, 6\}$ en cada uno de los primeros n lanzamientos del dado. Entonces claramente $A_n \supseteq A_{n+1}$ y $P(A_n) = 1/2^n$ para cualquier n en \mathbb{N} . Por lo tanto

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Entonces

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} 1/2^n = 0.$$

El evento $\bigcap_{n=1}^{\infty} A_n$ se interpreta como aquel conjunto de resultados en el que siempre se obtiene un número par en cada uno de los lanzamientos. Hemos demostrado que la probabilidad de tal evento es cero. En consecuencia la probabilidad de que eventualmente aparezca un número impar es uno. Observe que el argumento presentado funciona de la misma forma cuando el evento A es cualquier subconjunto propio de Ω distinto del vacío. Por ejemplo, si $A = \{1, 2, 3, 4, 5\}$, entonces la probabilidad de nunca obtener “6” es cero. Por lo tanto, con probabilidad uno, cada una de las caras del dado aparecerá eventualmente. Puede demostrarse además que cada una de las caras aparecerá una infinidad de veces con probabilidad uno. ■

1.4. Independencia de eventos

En esta sección se define el concepto importante de independencia de eventos. La independencia es un tema central en la teoría de la probabilidad, y uno de sus rasgos distintivos. De manera natural la independencia aparecerá con frecuencia a lo largo del texto a partir de ahora, y ayudará a

simplificar el cálculo de probabilidades. La definición matemática es la siguiente.

DEFINICIÓN. (INDEPENDENCIA DE DOS EVENTOS). Dos eventos A y B son independientes, y se escribe $A \perp B$, cuando

$$P(A \cap B) = P(A)P(B).$$

A menudo aceptar la hipótesis de que dos eventos son independientes es una cuestión de apreciación por parte del observador. La independencia puede interpretarse en el sentido de que la ocurrencia de uno de los eventos no proporciona información que modifique la probabilidad de ocurrencia del segundo evento. Contrario a alguna primera concepción intuitiva errónea, el hecho de que dos eventos sean independientes no implica que ellos sean ajenos. La proposición contraria tampoco es válida, dos eventos ajenos no necesariamente son independientes.

EJERCICIO. Demuestre que un evento es independiente consigo mismo si, y sólo si, su probabilidad es cero o uno. ▪

EJERCICIO. Demuestre que un evento que tiene probabilidad cero o uno, es independiente de cualquier otro evento, incluyendo él mismo. ▪

EJERCICIO. Demuestre que los eventos A y B son independientes si, y sólo si, ▪
 a) A y B lo son. b) A^c y B lo son. c) A y B^c lo son.

La definición de independencia puede extenderse a colecciones finitas e incluso infinitas de eventos del siguiente modo.

DEFINICIÓN. (INDEPENDENCIA DE VARIOS EVENTOS). Los eventos A_1, \dots, A_n son independientes si se cumplen todas y cada una de las siguientes condiciones:

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad i, j \text{ distintos.} \quad (1.1)$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} \quad (1.2)$$

$$\vdots$$

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdots P(A_n).$$

Más generalmente, una colección infinita de eventos es independiente si cualquier subcolección finita lo es.

Observe que de acuerdo a la definición anterior, se necesitan verificar o suponer varias condiciones para que n eventos sean independientes entre sí. De hecho el número total de igualdades a demostrar es $2^n - n - 1$. ¿Puede usted demostrar esta afirmación? En la siguiente sección haremos uso del siguiente resultado.

EJERCICIO. Demuestre que los eventos A_1, \dots, A_n son independientes si, y sólo si, los eventos A_1^c, \dots, A_n^c lo son. ■

Es posible además demostrar que la independencia dos a dos, igualdad (1.1) en la definición, no implica en general la independencia tres a tres, igualdad (1.2), ni viceversa.

EJERCICIO. Se lanza una moneda equilibrada tres veces. Defina los eventos $A = \text{“Se obtiene el mismo resultado en el 1er. y 2do. lanzamiento”}$.
 $B = \text{“Se obtiene el mismo resultado en el 2do. y 3er. lanzamiento”}$.
 $C = \text{“Se obtiene el mismo resultado en el 3er. y 1er. lanzamiento”}$.
 Demuestre que los eventos A, B y C son independientes dos a dos, pero no independientes en su conjunto. ■

EJERCICIO. Sean A y B eventos no independientes, y sea $C = \emptyset$. Demuestre que A, B y C son independientes tres a tres pero no son independientes dos

a dos. ▪

También se tiene la noción de independencia entre dos o más clases de eventos. La definición es la siguiente, como siempre se presupone un espacio de probabilidad (Ω, \mathcal{F}, P) dado.

DEFINICIÓN. (INDEPENDENCIA DE CLASES). Las clases no vacías de eventos $\mathcal{C}_1, \dots, \mathcal{C}_n$ son independientes si los eventos A_1, \dots, A_n lo son para cualesquiera A_i en \mathcal{C}_i , $i = 1, \dots, n$. Más generalmente, un conjunto infinito de clases no vacías de eventos es independiente si cualquier subconjunto finito lo es.

En particular, dos σ -álgebras \mathcal{F}_1 y \mathcal{F}_2 son *independientes* si para cada A en \mathcal{F}_1 y cada B en \mathcal{F}_2 se cumple que $P(A \cap B) = P(A)P(B)$. Análogamente para un número finito de σ -álgebras o bien un número infinito de ellas.

EJEMPLO. (EL PROBLEMA DEL MONO). Un mono escribe caracteres al azar en una máquina de escribir. ¿Cuál es la probabilidad de que eventualmente obtenga exactamente, y sin ningún error, las obras completas de Shakespeare?

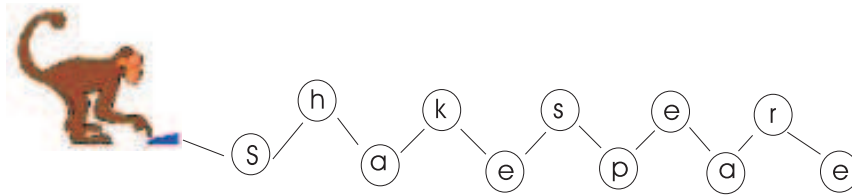


Figura 1.4: Mono escribiendo al azar.

Demostramos a continuación que la probabilidad de este raro evento es uno. Imagine entonces que un mono escribe caracteres al azar en una máquina de escribir, y que lo hace de manera continua generando una sucesión lineal de caracteres. Sea m el total de caracteres disponibles en una máquina de escribir, y sea N el total de caracteres de los que constan las obras comple-

tas de Shakespeare. Segmentamos el arreglo lineal de caracteres generados por el mono en bloques disjuntos de N caracteres, uno después de otro, y observamos si algún bloque contiene las obras de Shakespeare. Por ejemplo,

$$\underbrace{Xku \cdots aTs}_N \quad \underbrace{hwW \cdots pZq}_N \quad Ot \cdots$$

Para cada número natural k defina el evento A_k correspondiente a que el k -ésimo bloque contiene exactamente, y sin error alguno, las obras completas de Shakespeare. Observe que los eventos A_k son independientes pues los bloques no se sobreponen, además $P(A_k) = (1/m)^N = p$, o bien $P(A_k^c) = 1 - p$. Defina el evento B_k como $A_1^c \cap \cdots \cap A_k^c$, que indica la situación en la que el mono no obtiene éxito en los primeros k bloques. Observe que $B_{k+1} \subseteq B_k$, es decir la sucesión es decreciente, por lo tanto

$$\lim_{k \rightarrow \infty} B_k = \bigcap_{k=1}^{\infty} B_k,$$

en donde el evento $\bigcap_{k=1}^{\infty} B_k$ se interpreta como aquel en el que el mono nunca tiene éxito. Entonces, usando la propiedad de continuidad de las medidas de probabilidad para sucesiones decrecientes, se tiene que

$$P\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k \rightarrow \infty} P(B_k) = \lim_{k \rightarrow \infty} (1 - p)^k = 0.$$

Por lo tanto la probabilidad del evento complemento es uno, es decir, la probabilidad de que eventualmente el mono obtenga éxito es uno. Más adelante se presentarán otras formas de resolver este mismo problema usando el lema de Borel-Cantelli, y después usando la ley fuerte de los grandes números. En [25] aparece una estimación del tiempo promedio de espera para que el mono obtenga el primer éxito. \blacksquare

1.5. Lema de Borel-Cantelli

Concluimos este capítulo con el enunciado y demostración del famoso lema de Borel-Cantelli. El objetivo es demostrar este resultado y con ello poner

en práctica algunas propiedades de las medidas de probabilidad, aunque también lo usaremos para presentar un par de aplicaciones y para demostrar la ley fuerte de los grandes números en la última parte del curso.

PROPOSICIÓN. (LEMA DE BOREL-CANTELLI). Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos, y defina $A = \limsup_{n \rightarrow \infty} A_n$.

1. Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces $P(A) = 0$.
2. Si A_1, A_2, \dots son independientes y $\sum_{n=1}^{\infty} P(A_n) = \infty$, entonces $P(A) = 1$.

Demostración.

1. Para cada número natural n ,

$$P(A) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k).$$

Como $\sum_{n=1}^{\infty} P(A_n) < \infty$, el lado derecho tiende a cero cuando n tiende a infinito. Esto implica que $P(A) = 0$.

2. Es suficiente demostrar que para todo número natural n se cumple la igualdad $P(\bigcup_{k=n}^{\infty} A_k) = 1$, pues la intersección numerable de eventos

con probabilidad uno tiene probabilidad uno. Para cada $m > n$,

$$\begin{aligned}
 1 - P\left(\bigcup_{k=n}^{\infty} A_k\right) &\leq 1 - P\left(\bigcup_{k=n}^m A_k\right) \\
 &= P\left(\bigcap_{k=n}^m A_k^c\right) \\
 &= \prod_{k=n}^m [1 - P(A_k)] \\
 &\leq \exp\left(-\sum_{k=n}^m P(A_k)\right).
 \end{aligned}$$

Para obtener la última expresión se usa la desigualdad: $1 - x \leq e^{-x}$, válida para cualquier número real x . Como $\sum_{n=1}^{\infty} P(A_n) = \infty$, el lado derecho tiende a cero cuando m tiende a infinito. Por lo tanto $P\left(\bigcup_{k=n}^{\infty} A_k\right) = 1$ para cualquier valor de n y entonces $P(A) = 1$.

□

EJEMPLO. (EL PROBLEMA DEL MONO, NUEVAMENTE). El problema de encontrar la probabilidad de que un mono que escribe caracteres al azar en una máquina de escribir, eventualmente escriba las obras completas de Shakespeare, puede resolverse también usando el lema de Borel-Cantelli. Suponga que N es el total de caracteres de los que constan las obras completas de Shakespeare y considere nuevamente la división por bloques de longitud N :

$$\underbrace{x_1, \dots, x_N}_{\text{bloqueo 1}}, \underbrace{x_{N+1}, \dots, x_{2N}}_{\text{bloqueo 2}}, \dots$$

El evento A_k se define nuevamente como aquel en el que el mono tiene éxito en el k -ésimo bloque. Si nuevamente m denota el total de caracteres disponibles, entonces la probabilidad del evento A_k es $(1/m)^N$, y claramente la sucesión A_1, A_2, \dots constituye una sucesión de eventos independientes tales que $\sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^{\infty} (1/m)^N = \infty$. Entonces por la segunda parte del

lema de Borel-Cantelli, la probabilidad del límite superior de la sucesión A_k es uno. Ahora sólo hay que recordar que el evento $\limsup_{k \rightarrow \infty} A_k$ corresponde a aquel en el que una infinidad de eventos A_k ocurren. Es decir, con probabilidad uno, el mono tiene, no uno, sino ¡una infinidad de éxitos! .

EJERCICIO. Se lanza una moneda honesta una infinidad de veces. Use el lema de Borel-Cantelli para demostrar que la probabilidad de que cada cara aparezca una infinidad de veces es uno. ¿Importa que la moneda sea honesta? .

EJERCICIO. Sea x_1, \dots, x_n una sucesión de resultados consecutivos particular obtenida de lanzar una moneda n veces. Considere ahora el experimento de lanzar la moneda una infinidad de veces. Use el lema de Borel-Cantelli para calcular la probabilidad de que aparezca una infinidad de veces la sucesión particular mencionada. .



ANDREY NIKOLAEVICH KOLMOGOROV (RUSIA 1903–1987)

Creció bajo el amparo de su tía Vera Yakovlena, pues su madre murió en el parto y su padre fue exiliado. Trabajó un tiempo como conductor de trenes. En 1920 ingresó a la Universidad Estatal de Moscú, en donde además de matemáticas tomó cursos de metalurgia y sobre historia de Rusia. Aún siendo estudiante de licenciatura empezó a publicar trabajos de investigación graduándose en 1925. Terminó su doctorado en 1929, y para entonces ya tenía 18 publicaciones. Contribuyó brillantemente en varias áreas de las matemáticas como: análisis, probabilidad, procesos estocásticos, lógica, análisis funcional, geometría, topología, sistemas dinámicos, movimiento de los planetas, turbulencia, etc. Kolmogorov tenía particular interés en proveer de atención y educación especial a niños con habilidades sobresalientes. Recibió un sinnúmero de premios y reconocimientos de distintos países, y fue miembro de varias sociedades y academias científicas. Fuente: Archivo MacTutor, Universidad de St. Andrews.

1.6. Ejercicios

σ -álgebras

1. DEFINICIÓN ALTERNATIVA DE σ -ÁLGEBRA. Demuestre que \mathcal{F} es una σ -álgebra de subconjuntos de Ω si, y sólo si, satisface las siguientes propiedades:
 - a) $\emptyset \in \mathcal{F}$.
 - b) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
 - c) Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.
2. DEFINICIÓN ALTERNATIVA DE σ -ÁLGEBRA. Demuestre que \mathcal{F} es una σ -álgebra de subconjuntos de Ω si, y sólo si, satisface las siguientes propiedades:
 - a) $\Omega \in \mathcal{F}$.
 - b) $A, B \in \mathcal{F} \Rightarrow A - B \in \mathcal{F}$.
 - c) Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.
3. Sean A_1, \dots, A_n eventos de un espacio muestral Ω . Demuestre que el conjunto de elementos de Ω que pertenecen a exactamente k de estos eventos es un evento, $1 \leq k \leq n$.
4. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω . Demuestre que la colección $\mathcal{F}^c = \{F^c : F \in \mathcal{F}\}$ es una σ -álgebra. Compruebe que \mathcal{F}^c y \mathcal{F} coinciden.
5. Sea $\Omega = \{a, b, c, d\}$, y sean $A = \{a, b\}$ y $B = \{b, c\}$. Defina la colección $\mathcal{C} = \{A, B\}$. Claramente \mathcal{C} no es una σ -álgebra. Encuentre $\sigma(\mathcal{C})$.
6. Sea \mathcal{F} una σ -álgebra de subconjuntos de Ω y sea A un elemento de \mathcal{F} . Demuestre que la colección $\{A \cap F : F \in \mathcal{F}\}$ es una σ -álgebra de subconjuntos de A . Se usan los símbolos \mathcal{F}_A ó $A \cap \mathcal{F}$ para denotar a esta colección.

7. Sean Ω_1 y Ω_2 dos conjuntos arbitrarios, y sea $X : \Omega_1 \rightarrow \Omega_2$ una función en donde $(\Omega_2, \mathcal{F}_2)$ es un espacio medible. Demuestre que la siguiente colección es una σ -álgebra de subconjuntos de Ω_1 :

$$X^{-1}\mathcal{F}_2 = \{X^{-1}F : F \in \mathcal{F}_2\}.$$

8. ¿Es la diferencia de dos σ -álgebras una σ -álgebra? Demuestre o proporcione un contraejemplo.
9. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de Ω . Demuestre que $\mathcal{F}_1 \cup \mathcal{F}_2$ no necesariamente es una σ -álgebra. Para ello considere el espacio $\Omega = \{1, 2, 3\}$ y las σ -álgebras $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ y $\mathcal{F}_2 = \{\emptyset, \{1, 2\}, \{3\}, \Omega\}$.
10. Sean \mathcal{F}_1 y \mathcal{F}_2 dos σ -álgebras de subconjuntos de Ω tales que $\mathcal{F}_1 \subseteq \mathcal{F}_2$. Demuestre que $\mathcal{F}_1 \cup \mathcal{F}_2$ es una σ -álgebra.
11. Sea T un conjunto arbitrario distinto del vacío. Suponga que para cada t en T se tiene una σ -álgebra \mathcal{F}_t de subconjuntos de Ω . Demuestre con detalle que $\bigcap_{t \in T} \mathcal{F}_t$ es una σ -álgebra.
12. Sean $A, B \subseteq \Omega$ arbitrarios. Demuestre que la cardinalidad de $\sigma\{A, B\}$ es a lo sumo 16.
13. Sean $A, B \subseteq \Omega$ arbitrarios. Encuentre explícitamente todos los elementos de $\sigma\{A, B\}$. Por el ejercicio anterior, el total de elementos en $\sigma\{A, B\}$ es, en el caso más general, 16.
14. Sea $\{A_1, \dots, A_n\}$ una partición finita de Ω , es decir, la unión de todos estos conjuntos es Ω , ninguno de ellos es vacío y la intersección de cualesquiera dos de ellos es vacía. Demuestre que la cardinalidad de $\sigma\{A_1, \dots, A_n\}$ es 2^n .
15. Demuestre que toda σ -álgebra de un espacio muestral finito contiene un número par de elementos.
16. Sea $\{A, B, C\}$ una partición de Ω . Encuentre explícitamente los ocho elementos de $\sigma\{A, B, C\}$.

17. Sea \mathcal{C} una colección de subconjuntos de Ω . Diga falso o verdadero justificando en cada caso: $\mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq 2^\Omega$.
18. Demuestre que 2^Ω es una σ -álgebra de subconjuntos de Ω y que no existe una σ -álgebra de subconjuntos de Ω que sea más grande.
19. Sea Ω un conjunto, \mathcal{F} una σ -álgebra de subconjuntos de Ω y sea A un evento cualquiera. De cada una de las dos expresiones siguientes determine la que es notacionalmente correcta. Explique su respuesta.
- $\Omega \in \mathcal{F}$ ó $\Omega \subseteq \mathcal{F}$.
 - $A \in \Omega$ ó $A \subseteq \Omega$.
 - $\emptyset \in \mathcal{F}$ ó $\emptyset \subseteq \mathcal{F}$.
 - $A \in \mathcal{F}$ ó $A \subseteq \mathcal{F}$.

σ -álgebras, álgebras y semiálgebras

20. DEFINICIÓN ALTERNATIVA DE ÁLGEBRA. Demuestre que \mathcal{F} es una álgebra de subconjuntos de Ω si, y sólo si, cumple las siguientes condiciones:
- $\Omega \in \mathcal{F}$.
 - Si $A, B \in \mathcal{F}$, entonces $A - B \in \mathcal{F}$.

21. Demuestre que

$$\mathcal{F} \text{ es } \sigma\text{-álgebra} \Rightarrow \mathcal{F} \text{ es álgebra} \Rightarrow \mathcal{F} \text{ es semiálgebra.}$$

22. ÁLGEBRA $\not\Rightarrow$ σ -ÁLGEBRA. Sea $\Omega = (0, 1]$ y defina la colección \mathcal{F} de subconjuntos de la forma

$$\bigcup_{i=1}^n (a_i, b_i],$$

en donde $(a_i, b_i] \subseteq (0, 1]$ con $(a_i, b_i] \cap (a_j, b_j] = \emptyset$ para $i \neq j$ y $n \in \mathbb{N}$. Demuestre que \mathcal{F} es una álgebra pero no una σ -álgebra.

23. Mediante un contraejemplo demuestre que no toda semiálgebra es una álgebra.

Conjuntos de Borel

24. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(a, b] : a \leq b\}$.
25. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{[a, b) : a \leq b\}$.
26. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(a, \infty) : a \in \mathbb{R}\}$.
27. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, a) : a \in \mathbb{R}\}$.
28. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, b] : b \in \mathbb{R}\}$.
29. Demuestre que $\mathcal{B}(\mathbb{R}) = \sigma\{(a, \infty) : a \in \mathbb{R}\}$.
30. Sea $A \in \mathcal{B}(\mathbb{R})$. Demuestre que $\mathcal{B}(A)$ es efectivamente una σ -álgebra de subconjuntos de A .
31. Diga falso o verdadero. Justifique su respuesta.
- $\sigma\left\{\left(\frac{1}{n+1}, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \mathcal{B}(0, 1]$.
 - $\sigma\left\{\left(0, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \mathcal{B}(0, 1]$.
 - $\sigma\left\{\left(\frac{1}{n+1}, \frac{1}{n}\right] : n \in \mathbb{N}\right\} = \sigma\left\{\left(0, \frac{1}{n}\right] : n \in \mathbb{N}\right\}$.
32. Demuestre que $\mathcal{B}(\mathbb{R}^2) = \sigma\{[a, b] \times [c, d] : a \leq b, c \leq d\}$.
33. Demuestre que $\mathcal{B}(\mathbb{R}^2) = \sigma\{(-\infty, a) \times (-\infty, b) : a, b \in \mathbb{R}\}$.
34. Demuestre que $\mathcal{B}(\mathbb{R}^2) = \sigma\{(a, \infty) \times (b, \infty) : a, b \in \mathbb{R}\}$.

Sucesiones de eventos

35. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que
- $\limsup_{n \rightarrow \infty} A_n$ es un evento.

- b) $\liminf_{n \rightarrow \infty} A_n$ es un evento.
 c) $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$.

36. Demuestre que el evento

- a) $\limsup_{n \rightarrow \infty} A_n$ coincide con el conjunto
 $\{\omega \in A_n \text{ para una infinidad de valores de } n\}$.
 b) $\liminf_{n \rightarrow \infty} A_n$ coincide con el conjunto
 $\{\omega \in A_n \text{ para toda } n \text{ excepto un número finito de ellas}\}$.

37. Suponga $A_n \subseteq B_n$ para cada n en \mathbb{N} . Demuestre o proporcione un contraejemplo.

- a) $\limsup_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} B_n$.
 b) $\liminf_{n \rightarrow \infty} A_n \subseteq \liminf_{n \rightarrow \infty} B_n$.
 c) $\limsup_{n \rightarrow \infty} A_n \subseteq \liminf_{n \rightarrow \infty} B_n$.

38. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que

- a) $(\liminf_{n \rightarrow \infty} A_n)^c = \limsup_{n \rightarrow \infty} A_n^c$.
 b) $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} A_n^c$.
 c) $P(\liminf_{n \rightarrow \infty} A_n) = 1 - P(\limsup_{n \rightarrow \infty} A_n^c)$.
 d) $P(\limsup_{n \rightarrow \infty} A_n) = 1 - P(\liminf_{n \rightarrow \infty} A_n^c)$.

39. Sea $\{A_n : n \in \mathbb{N}\}$ una sucesión de eventos. Demuestre que

- a) $\lim_{n \rightarrow \infty} A_n = A \Leftrightarrow \lim_{n \rightarrow \infty} A_n^c = A^c$.
 b) $\lim_{n \rightarrow \infty} A_n = A \Leftrightarrow \lim_{n \rightarrow \infty} 1_{A_n} = 1_A$.

El símbolo 1_A denota la función indicadora del conjunto A . Véase el apéndice al final del texto para la definición y algunas propiedades de esta función.

40. Sea $\{a_n : n \in \mathbb{N}\}$ una sucesión de números no negativos convergente al número $a \geq 0$. Sea $A_n = [0, a_n]$. Calcule $\liminf_{n \rightarrow \infty} A_n$ y $\limsup_{n \rightarrow \infty} A_n$.
41. Determine si cada una de las siguientes sucesiones de conjuntos es convergente.
- $A_n = (1/n, 2 + (-1)^n) \subseteq \mathbb{R}$.
 - $A_n = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq (1 + 1/n)^n\}$.
 - $A_n = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 2 + \text{sen}(n\pi/2)\}$.
42. Demuestre que las siguientes sucesiones de eventos no son convergentes.
- $A_n = \emptyset$ si n es impar, y $A_n = \Omega$ si n es par.
 - $A_n = (0, 1 + (-1)^n) \subseteq \mathbb{R}$.
43. Suponga que $\lim_{n \rightarrow \infty} A_n = A$, y $\lim_{n \rightarrow \infty} B_n = B$. Determine si la siguiente sucesión es convergente.

$$C_n = \begin{cases} A_n & \text{si } n \text{ es impar,} \\ B_n & \text{si } n \text{ es par.} \end{cases}$$

44. Encuentre condiciones sobre los eventos A y B para que la siguiente sucesión de eventos sea convergente.

$$A_n = \begin{cases} A & \text{si } n \text{ es impar,} \\ B & \text{si } n \text{ es par.} \end{cases}$$

45. Suponga que $\lim_{n \rightarrow \infty} A_n = A$. Demuestre que para cualquier evento B ,
- $\lim_{n \rightarrow \infty} (A_n \cap B) = A \cap B$.
 - $\lim_{n \rightarrow \infty} (A_n \cup B) = A \cup B$.
 - $\lim_{n \rightarrow \infty} (A_n - B) = A - B$.
 - $\lim_{n \rightarrow \infty} (A_n \triangle B) = A \triangle B$.

46. Suponga que $\lim_{n \rightarrow \infty} A_n = A$ y $\lim_{n \rightarrow \infty} B_n = B$. Diga falso o verdadero. Demuestre en cada caso.

a) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \cap B_m) = A \cap B$.

b) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \cup B_m) = A \cup B$.

c) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n - B_m) = A - B$.

d) $\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} (A_n \Delta B_m) = A \Delta B$.

47. Suponga que $\lim_{n \rightarrow \infty} A_n = A$ y $\lim_{n \rightarrow \infty} B_n = B$. Diga falso o verdadero. Demuestre en cada caso.

a) $\lim_{n \rightarrow \infty} (A_n \cap B_n) = A \cap B$.

b) $\lim_{n \rightarrow \infty} (A_n \cup B_n) = A \cup B$.

c) $\lim_{n \rightarrow \infty} (A_n - B_n) = A - B$.

d) $\lim_{n \rightarrow \infty} (A_n \Delta B_n) = A \Delta B$.

Medidas de probabilidad

48. Determine completamente un espacio de probabilidad (Ω, \mathcal{F}, P) para el experimento aleatorio de

- lanzar una moneda equilibrada.
- lanzar un dado equilibrado.
- escoger al azar un número real dentro del intervalo unitario $[0, 1]$.
- extraer dos bolas de una urna en donde hay dos bolas blancas y dos negras.
- lanzar una moneda honesta repetidas veces hasta que hayan aparecido ambas caras.

49. MEDIDA DE PROBABILIDAD DISCRETA. Sea $\{x_n : n \in \mathbb{N}\}$ una sucesión de números reales y sea $\{a_n : n \in \mathbb{N}\}$ otra sucesión de números reales no negativos tal que $\sum_{n=1}^{\infty} a_n = 1$. Demuestre que la función $P : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ definida de la siguiente forma es una medida de probabilidad.

$$P(A) = \sum_{n=1}^{\infty} a_n 1_{\{n : x_n \in A\}}(n).$$

50. Sean P y Q dos medidas de probabilidad definidas sobre una misma σ -álgebra. Demuestre que $\alpha P + (1 - \alpha)Q$ es una medida de probabilidad para cada α en $[0, 1]$.
51. Sea P una medida de probabilidad. Determine si las siguientes funciones también son medidas de probabilidad:

$$\begin{array}{lll} a) & 1 - P. & c) & P^2. & e) & 4P(1 - P). \\ b) & (1 + P)/2. & d) & |P|. & f) & \sqrt{P}. \end{array}$$

52. Determine si las siguientes funciones son medidas de probabilidad.

$$\begin{array}{ll} a) & P(\Omega) = 1 \text{ y } P(A) = 0 \text{ para cualquier otro evento } A. \\ b) & P(\emptyset) = 0 \text{ y } P(A) = 1 \text{ para cualquier otro evento } A. \end{array}$$

53. Considere el espacio medible $(\mathbb{N}, 2^{\mathbb{N}})$. Demuestre en cada caso que P es una medida de probabilidad. Para cada $A \in 2^{\mathbb{N}}$ defina:

$$a) P(A) = \sum_{n \in A} 2/3^n.$$

$$b) P(A) = \sum_{n \in A} 1/2^n.$$

54. Sea $\Omega = \{1, \dots, n\}$, y considere el espacio medible $(\Omega, 2^{\Omega})$. Investigue en cada caso si P es una medida de probabilidad. Para cada $A \in 2^{\Omega}$ defina:

$$a) P(A) = \sum_{k \in A} \frac{2k}{n(n+1)}.$$

$$b) P(A) = \prod_{k \in A} \left(1 - \frac{1}{k}\right).$$

55. Considere el espacio medible $((0, 1), \mathcal{B}(0, 1))$. Demuestre en cada caso que P es una medida de probabilidad. Para cada $A \in \mathcal{B}(0, 1)$ defina:

$$a) P(A) = \int_A 2x \, dx.$$

$$b) P(A) = \int_A \frac{3}{2} \sqrt{x} \, dx.$$

56. **PROBABILIDAD CONDICIONAL.** Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea B un evento con probabilidad estrictamente positiva. Demuestre que la probabilidad condicional definida para cada A en \mathcal{F} como sigue: $P(A|B) = P(A \cap B)/P(B)$, es una medida de probabilidad. En consecuencia, toda propiedad válida para una medida de probabilidad es también válida para la probabilidad condicional.

57. Sea P una medida de probabilidad, y sean $P_1(\cdot) = P(\cdot|B)$ y $P_2(\cdot) = P_1(\cdot|C)$, en donde $P(B) > 0$ y $P(C) > 0$. Demuestre que para cualquier evento A , $P_2(A) = P(A|B \cap C)$.

58. Demuestre que $P(A|B) \geq 1 - P(A^c)/P(B)$, en donde $P(B) > 0$.

59. Sea P una medida de probabilidad definida sobre la σ -álgebra \mathcal{F} . Demuestre que la colección $\{A \in \mathcal{F} : P(A) = 0 \text{ ó } P(A) = 1\}$ es una sub σ -álgebra de \mathcal{F} .

Propiedades elementales

60. Demuestre que $P(\emptyset) = 0$, sin usar $P(\Omega) = 1$.

61. Demuestre que $P(A \cap B) - P(A)P(B) = P(A^c)P(B) - P(A^c \cap B)$.

62. Demuestre que

$$P(A \cap B) \leq \min\{P(A), P(B)\} \leq P(A) \leq \max\{P(A), P(B)\} \leq P(A \cup B).$$

63. Demuestre que

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$

64. Demuestre que

$$P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C).$$

65. Demuestre que

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P(A_1) + P(A_1^c \cap A_2) + P(A_1^c \cap A_2^c \cap A_3) + \cdots \\ &\quad + P(A_1^c \cap \cdots \cap A_{n-1}^c \cap A_n) + \cdots \end{aligned}$$

66. FÓRMULA DE INCLUSIÓN Y EXCLUSIÓN. Demuestre que

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ &\quad - \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). \end{aligned}$$

67. Demuestre que

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i \cup A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cup A_j \cup A_k) \\ &\quad - \cdots + (-1)^{n+1} P(A_1 \cup \cdots \cup A_n). \end{aligned}$$

68. Demuestre que $P\left(\bigcap_{k=1}^n A_k\right) \geq 1 - \sum_{k=1}^n P(A_k^c)$.

69. Demuestre que

$$0 \leq P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B) \leq 2.$$

70. Diga falso o verdadero. Demuestre en cada caso.

- a) $P(B - A) = P(B) - P(A)$.
- b) $P(A \cup B) = P(A - B) + P(B - A)$.
- c) $P(A) > 0 \Rightarrow P(A \cup B) > 0$.
- d) $P(A) > 0 \Rightarrow P(A \cap B) > 0$.
- e) $P(A) < 1 \Rightarrow P(A \cup B) < 1$.
- f) $P(A) < 1 \Rightarrow P(A \cap B) < 1$.

71. Diga falso o verdadero. Demuestre en cada caso.

- a) $P(A) = 0 \Rightarrow P(A \cup B) = 0$.
- b) $P(A) = 0 \Rightarrow P(A \cap B) = 0$.
- c) $P(A \cup B) = 0 \Rightarrow P(A) = 0$.
- d) $P(A \cap B) = 0 \Rightarrow P(A) = 0$.
- e) $P(A) = 1 \Rightarrow P(A \cup B) = 1$.
- f) $P(A) = 1 \Rightarrow P(A \cap B) = 1$.
- g) $P(A \cup B) = 1 \Rightarrow P(A) = 1$.
- h) $P(A \cap B) = 1 \Rightarrow P(A) = 1$.

72. Diga falso o verdadero. Demuestre en cada caso.

- a) $P(A \cap B) \geq P(A) - P(B^c)$.
- b) $P(A - B) = P(A) - P(A \cap B)$.
- c) $P(A \cap B) \leq P(A)P(B)$.

- d) $P(A \cup B) \leq P(A) + P(B)$.
 e) $P(A|B) \leq P(A)$.
 f) $P(A|B) \geq P(A) \Rightarrow P(B|A) \geq P(B)$.
73. Se lanza una moneda tantas veces como indica un dado previamente lanzado. Tanto la moneda como el dado están equilibrados. Calcule la probabilidad de que:
- se obtengan ambas caras de la moneda igual número de veces.
 - se obtenga una misma cara siempre.
74. En una primera caja se encuentran dos canicas blancas y tres negras, en una segunda caja hay tres blancas y cinco negras, y en una tercera caja hay dos blancas y una negra. De la primera caja se extrae al azar una canica y se deposita en la segunda caja, después se extrae nuevamente al azar una canica de la segunda caja y se deposita en la tercera caja. Después de este proceso se obtiene al azar una canica de la tercera caja, encuentre la probabilidad de que ésta sea blanca.
75. Un dado equilibrado se lanza tres veces consecutivas, y resulta que la suma de los tres números obtenidos es 11. Encuentre la probabilidad de que en el primer lanzamiento se haya obtenido un 5.
76. Una primera caja contiene tres canicas blancas y dos negras. Una segunda caja contiene dos canicas blancas y cuatro negras. Se escoge una caja al azar y se extrae una canica. Únicamente se conoce que la canica obtenida es blanca, encuentre la probabilidad de que ésta haya sido obtenida de la primera caja.
77. REGLA DEL PRODUCTO. Demuestre que
- $$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$
78. DESIGUALDAD DE BONFERRONI. Demuestre que

$$P\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j).$$

79. DESIGUALDAD DE KOUNIAS. Demuestre que

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \min_j \left\{ \sum_{i=1}^n P(A_i) - \sum_{\substack{i=1 \\ i \neq j}}^n P(A_i \cap A_j) \right\}.$$

Continuidad

80. Se lanza una moneda honesta una infinidad de veces. Demuestre que la probabilidad de que eventualmente cada una de las dos caras aparezca es uno.
81. Se lanza un dado equilibrado una infinidad de veces. Demuestre que la probabilidad de que eventualmente cada una de las seis caras aparezca es uno.
82. Sea A un evento con probabilidad estrictamente positiva. Demuestre que si se efectúa una infinidad de ensayos independientes del experimento aleatorio, la probabilidad de que nunca ocurra el evento A es cero.

Independencia de eventos

83. Diga falso o verdadero. Demuestre o proporcione un contraejemplo.
 a) $A \perp A$. b) $A \perp A^c$. c) $A \perp \emptyset$. d) $A \perp \Omega$.
84. ¿Es la independencia de dos eventos una relación de equivalencia?
85. Mediante un contraejemplo demuestre que si A y B son independientes, entonces no necesariamente son ajenos. Demuestre también que si A y B son ajenos, entonces tampoco se sigue necesariamente que estos eventos sean independientes.

86. Sean A_1, \dots, A_n independientes. Demuestre que

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - \prod_{k=1}^n [1 - P(A_k)].$$

87. Sea A_1, A_2, \dots una sucesión infinita de eventos. Defina

$$B_n = \bigcup_{k=n}^{\infty} A_k \quad \text{y} \quad C_n = \bigcap_{k=n}^{\infty} A_k.$$

Demuestre que si B_n y C_n son independientes para cada n , entonces los eventos límite superior y límite inferior de la sucesión A_n también son independientes. En particular, cuando la sucesión A_n converge al evento A , entonces A tiene probabilidad cero o uno.

88. Sean A y B independientes. Demuestre que $\sigma\{A\}$ y $\sigma\{B\}$ son independientes.

Lema de Borel-Cantelli

89. Se lanza un dado equilibrado una infinidad de veces. Demuestre que con probabilidad uno cada una de las seis caras aparece una infinidad de veces.

90. Sea A un evento con probabilidad positiva. Use el lema de Borel-Cantelli para demostrar que si se efectúa una infinidad de ensayos independientes del experimento aleatorio, la probabilidad de que ocurra una infinidad de veces el evento A , es uno.

CAPÍTULO 2

Variables aleatorias

En este capítulo se estudian los conceptos de variable aleatoria, función de distribución, función de densidad y esperanza. Se estudian también algunas distribuciones de probabilidad de variables aleatorias discretas y continuas particulares. A partir de ahora y en el resto del curso consideraremos como elemento base un espacio de probabilidad (Ω, \mathcal{F}, P) .

2.1. Variables aleatorias

Una variable aleatoria es una función del espacio muestral en el conjunto de números reales que además satisface cierta condición de medibilidad. Representa una traducción de cada uno de los resultados del espacio muestral en números reales. Mediante una variable aleatoria uno puede considerar que el posible experimento aleatorio en cuestión no produce como resultados elementos de Ω sino números reales. El concepto de variable aleatoria es fundamental en la teoría de la probabilidad, y una vez que enunciemos su definición, el término aparecerá con mucha frecuencia a lo largo del curso.

DEFINICIÓN. (VARIABLE ALEATORIA). Una variable aleatoria real es una función $X : \Omega \rightarrow \mathbb{R}$ tal que para cualquier conjunto Boreliano B , se cumple que el conjunto $X^{-1}B$ es un elemento de \mathcal{F} .

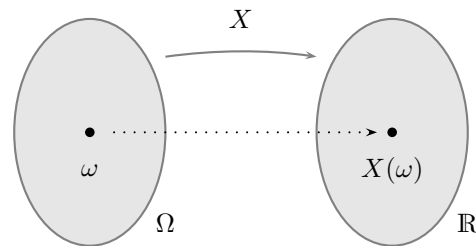


Figura 2.1: Una variable aleatoria es una función medible de Ω en \mathbb{R} .

Gráficamente una variable aleatoria puede representarse como se muestra en la Figura 2.1. Esto es, una variable aleatoria (a veces se escribe simplemente v.a.) es una función de Ω en \mathbb{R} tal que la imagen inversa de cualquier conjunto Boreliano es un elemento de la σ -álgebra del espacio de probabilidad. Esta condición se conoce como *medibilidad* en teoría de la medida, y se dice entonces que dicha función es *medible* respecto de las σ -álgebras \mathcal{F} y $\mathcal{B}(\mathbb{R})$. En un apéndice al final del texto aparece una sección que contiene una discusión breve del concepto de imagen inversa de una función, que para el caso de variables aleatorias puede ilustrarse gráficamente como se indica en la Figura 2.2.

Explicamos a continuación la razón técnica por la cual se le pide a una función $X : \Omega \rightarrow \mathbb{R}$ que cumpla la condición de medibilidad. Recordemos que P es una medida de probabilidad definida sobre el espacio medible (Ω, \mathcal{F}) . Si X es una variable aleatoria, entonces podemos trasladar la medida de probabilidad P al espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ del siguiente modo: Si B es un conjunto Boreliano definimos $P_X(B) = P(X^{-1}B)$, lo cual es posible pues el conjunto $X^{-1}B$ es un elemento de \mathcal{F} , dominio de definición de P . La función $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ resulta ser una medida de probabilidad, y se le llama por tanto la *medida de probabilidad inducida* por la variable aleatoria.

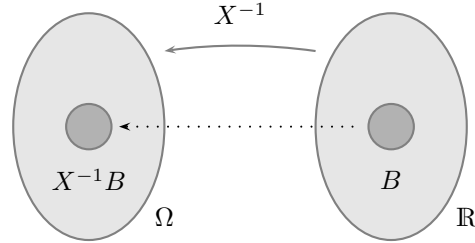


Figura 2.2: La imagen inversa de un conjunto de Borel.

Se le conoce también con el nombre de distribución o ley de probabilidad de X . A menudo se le denota por $\mathcal{L}(X)$. De este modo se construye el espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

Si B es un conjunto Boreliano, se usan los símbolos $X^{-1}B$ y $(X \in B)$ para denotar el conjunto $\{\omega \in \Omega : X(\omega) \in B\}$. Por ejemplo, el conjunto $\{\omega \in \Omega : X(\omega) \in [0, \infty)\}$ puede ser denotado por $X^{-1}[0, \infty)$ o $(X \in [0, \infty))$, o simplemente por $(X \geq 0)$, incluyendo los paréntesis. Veamos otro ejemplo, si (a, b) es un intervalo de la recta real, se puede usar el símbolo $X^{-1}(a, b)$, o $(X \in (a, b))$, o bien $(a < X < b)$ para denotar el conjunto $\{\omega \in \Omega : X(\omega) \in (a, b)\}$. Para hacer la escritura más corta, a menudo se omite el argumento ω de una variable X y se omite también el término variable aleatoria para X suponiendo, en la mayoría de las veces, que lo es.

Para comprobar que una función $X : \Omega \rightarrow \mathbb{R}$ es realmente una variable aleatoria, la definición requiere verificar la condición $X^{-1}B \in \mathcal{F}$ para cualquier conjunto Boreliano B . En muy pocos casos tal condición puede comprobarse de manera tan directa. La siguiente proposición establece que no es necesario demostrar la condición de medibilidad para cualquier conjunto Boreliano B , sino que es suficiente tomar intervalos de la forma $(-\infty, x]$, para cada x en \mathbb{R} . Este resultado, como uno puede imaginar, es de suma utilidad y lo usaremos con frecuencia en el resto del capítulo.

PROPOSICIÓN. Una función $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria si, y sólo si, para cada x en \mathbb{R} se cumple que $(X \leq x) \in \mathcal{F}$.

Demostración.

- (\Rightarrow) Si X es variable aleatoria, entonces claramente se cumple que para cualquier número real x el conjunto $(X \leq x)$ es un elemento de \mathcal{F} .
- (\Leftarrow) Ahora suponga que para cada real x , el conjunto $(X \leq x)$ es un elemento de \mathcal{F} . Sean \mathcal{B} y \mathcal{C} las colecciones

$$\begin{aligned}\mathcal{B} &= \{B \in \mathcal{B}(\mathbb{R}) : X^{-1}B \in \mathcal{F}\}, \\ \text{y } \mathcal{C} &= \{(-\infty, x] : x \in \mathbb{R}\}.\end{aligned}$$

Entonces claramente $\mathcal{C} \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R})$. La primera contención es por hipótesis, y la segunda es por definición de la colección \mathcal{B} . Suponga por un momento que \mathcal{B} es una σ -álgebra de subconjuntos de \mathbb{R} . Entonces \mathcal{B} es una σ -álgebra que contiene a \mathcal{C} . Por lo tanto $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B}$. Esto implica que $\mathcal{B} = \mathcal{B}(\mathbb{R})$, y entonces X es variable aleatoria. Resta entonces hacer ver que \mathcal{B} es efectivamente una σ -álgebra.

- a) Primeramente tenemos que $\mathbb{R} \in \mathcal{B}$, pues $\mathbb{R} \in \mathcal{B}(\mathbb{R})$ y $X^{-1}\mathbb{R} = \Omega \in \mathcal{F}$.
- b) Sea $B \in \mathcal{B}$. Entonces $B \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B \in \mathcal{F}$. Por lo tanto $B^c \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B^c = (X^{-1}B)^c \in \mathcal{F}$. Es decir, $B^c \in \mathcal{B}$.
- c) Sea B_1, B_2, \dots una sucesión en \mathcal{B} . Es decir, para cada número natural n , $B_n \in \mathcal{B}(\mathbb{R})$ y $X^{-1}B_n \in \mathcal{F}$. Entonces $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}(\mathbb{R})$ y $\bigcup_{n=1}^{\infty} X^{-1}B_n = X^{-1} \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$. Es decir, $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

□

Además de la condición anterior para demostrar que una función es variable aleatoria, existen otras condiciones igualmente equivalentes y útiles.

Por ejemplo, X es variable aleatoria si para cada x en \mathbb{R} , $(X < x) \in \mathcal{F}$, o $(X > x) \in \mathcal{F}$, o $(X \geq x) \in \mathcal{F}$. Cualquiera de estas condiciones es necesaria y suficiente para que X sea variable aleatoria. También es equivalente la condición $(a < X < b) \in \mathcal{F}$, para cualquier intervalo (a, b) de \mathbb{R} . La demostración de todas estas aseveraciones es completamente análoga al caso demostrado arriba y se pide desarrollar los detalles en la sección de ejercicios.

Considere ahora los espacios medibles (Ω, \mathcal{F}) y $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si X es una función de Ω en \mathbb{R} , entonces se denota por $\sigma(X)$ a la mínima σ -álgebra de subconjuntos de Ω respecto de la cual X es variable aleatoria.

DEFINICIÓN. $\sigma(X) = \{ X^{-1}B : B \in \mathcal{B}(\mathbb{R}) \}$.

Es sencillo probar que tal colección de imágenes inversas es efectivamente una σ -álgebra, y claramente X es variable aleatoria si, y sólo si, $\sigma(X) \subseteq \mathcal{F}$. En particular, se dice que una función $g : \mathbb{R} \rightarrow \mathbb{R}$ es *Borel medible* si $g^{-1}B \in \mathcal{B}(\mathbb{R})$, para cada B en $\mathcal{B}(\mathbb{R})$.

EJERCICIO. Diga falso o verdadero. Demuestre en cada caso.

a) $\sigma(X) = \sigma(X^2)$. b) $\sigma(X) = \sigma(-X)$. c) $\sigma(X) = \sigma(2X)$. .

A continuación se demuestra que algunas operaciones básicas entre variables aleatorias producen nuevas variables aleatorias. Suponga entonces que (Ω, \mathcal{F}, P) es un espacio de probabilidad dado. Todas las variables aleatorias que se consideran a continuación están definidas sobre este mismo espacio de probabilidad.

PROPOSICIÓN. La función constante $X = c$ es una variable aleatoria.

Demostración. Sea B un elemento cualquiera de $\mathcal{B}(\mathbb{R})$. Para la función constante $X = c$ se tiene que $X^{-1}B = \Omega$ si $c \in B$, y $X^{-1}B = \emptyset$ si $c \notin B$. En ambos casos el conjunto $X^{-1}B$ es un elemento de \mathcal{F} , por lo tanto X es

variable aleatoria. □

PROPOSICIÓN. Si X es variable aleatoria y c es una constante, entonces cX también es variable aleatoria.

Demostración. Comprobaremos que para cada número real x , la imagen inversa del conjunto $(-\infty, x]$, bajo la función cX , es un elemento de \mathcal{F} . Tenemos tres casos: Si $c > 0$, entonces el conjunto $(cX \leq x) = (X \leq x/c)$ es un elemento de \mathcal{F} , pues X es v.a. Si $c < 0$, entonces nuevamente el conjunto $(cX \leq x) = (X \geq x/c)$ es un elemento de \mathcal{F} pues X es v.a. Finalmente si $c = 0$, entonces es claro que cX es la constante cero que es v.a. por la proposición anterior. □

PROPOSICIÓN. Si X y Y son v.a.s, entonces $X + Y$ es variable aleatoria.

Demostración. Probaremos que para cada número real x , el conjunto $(X + Y > x)$ es un elemento de \mathcal{F} . Para ello usaremos la igualdad

$$(X + Y > x) = \bigcup_{r \in \mathbb{Q}} (X > r) \cap (Y > x - r). \quad (2.1)$$

Es claro que a partir de esta igualdad se concluye que el conjunto $(X + Y > x)$ es un elemento de \mathcal{F} , pues tanto X como Y son variables aleatorias, y la operación de unión involucrada es numerable. Resta entonces demostrar (2.1).

(\subseteq) Sea ω en Ω tal que $X(\omega) + Y(\omega) > x$. Entonces $X(\omega) > x - Y(\omega)$. Como los números racionales son un conjunto denso en \mathbb{R} , tenemos que existe un número racional r tal que $X(\omega) > r > x - Y(\omega)$. Por lo tanto $X(\omega) > r$ y $Y(\omega) > x - r$. De aquí se desprende que ω es un elemento del lado derecho.

- (\supseteq) Sea ahora ω un elemento de $\bigcup_{r \in \mathbb{Q}} (X > r) \cap (Y > x - r)$. Entonces existe un número racional r_0 tal que $X(\omega) > r_0$ y $Y(\omega) > x - r_0$. Sumando obtenemos $X(\omega) + Y(\omega) > x$, y por lo tanto ω es también un elemento del lado izquierdo.

□

PROPOSICIÓN. Si X y Y son v.a.s, entonces XY es variable aleatoria.

Demostración. Suponga primero el caso particular $X = Y$. Entonces necesitamos probar que para todo número real x , el conjunto $(X^2 \leq x)$ es un elemento de \mathcal{F} . Pero esto es cierto pues $(X^2 \leq x) = \emptyset$ si $x < 0$, y $(X^2 \leq x) = (-\sqrt{x} \leq X \leq \sqrt{x})$ si $x \geq 0$. En ambos casos, el conjunto $(X^2 \leq x)$ es un elemento de \mathcal{F} . Para el caso general, $X \neq Y$, usamos la fórmula $XY = ((X + Y)^2 - (X - Y)^2)/4$. Por lo demostrado antes, el producto XY es efectivamente una variable aleatoria. □

Como consecuencia se cumple que si multiplicamos X por sí misma n veces, entonces X^n es variable aleatoria. Por lo tanto toda función polinomial de una variable aleatoria es también variable aleatoria.

PROPOSICIÓN. Sean X y Y v.a.s con $Y \neq 0$. Entonces X/Y es variable aleatoria.

Demostración. Como el producto de variables aleatorias es nuevamente una variable aleatoria, es suficiente demostrar que $1/Y$ es variable aleatoria. Para

cualquier número real $y > 0$ tenemos que

$$\begin{aligned} \left(\frac{1}{Y} \leq y\right) &= \left(\frac{1}{Y} \leq y, Y > 0\right) \cup \left(\frac{1}{Y} \leq y, Y < 0\right) \\ &= \left(Y \geq \frac{1}{y}, Y > 0\right) \cup \left(Y \leq \frac{1}{y}, Y < 0\right) \\ &= \left(Y \geq \frac{1}{y}\right) \cup \left(Y < 0\right), \end{aligned}$$

que es un elemento de \mathcal{F} puesto que Y es variable aleatoria. Por otro lado, si $y < 0$ tenemos que

$$\begin{aligned} \left(\frac{1}{Y} \leq y\right) &= \left(\frac{1}{Y} \leq y, Y > 0\right) \cup \left(\frac{1}{Y} \leq y, Y < 0\right) \\ &= \left(Y \leq \frac{1}{y}, Y > 0\right) \cup \left(Y \geq \frac{1}{y}, Y < 0\right) \\ &= \emptyset \cup \left(Y \geq \frac{1}{y}, Y < 0\right) \\ &= \left(\frac{1}{y} \leq Y < 0\right). \end{aligned}$$

Nuevamente vemos que este conjunto es un elemento de \mathcal{F} , puesto que Y es v.a. Finalmente cuando $y = 0$ obtenemos una vez mas un elemento de \mathcal{F} pues

$$\begin{aligned} \left(\frac{1}{Y} \leq 0\right) &= \left(\frac{1}{Y} \leq 0, Y > 0\right) \cup \left(\frac{1}{Y} \leq 0, Y < 0\right) \\ &= \emptyset \cup \left(Y < 0\right) \\ &= \left(Y < 0\right). \end{aligned}$$

□

PROPOSICIÓN. Si X y Y son variables aleatorias, entonces $\max\{X, Y\}$ y $\min\{X, Y\}$ también lo son.

Demostración. Para cualquier número real x ,

$$(\max\{X, Y\} \leq x) = (X \leq x, Y \leq x) = (X \leq x) \cap (Y \leq x).$$

Análogamente,

$$(\min\{X, Y\} \geq x) = (X \geq x, Y \geq x) = (X \geq x) \cap (Y \geq x).$$

□

Como consecuencia se obtiene que tanto $X^+ = \max\{0, X\}$ como $X^- = -\min\{0, X\}$ son variables aleatorias.

PROPOSICIÓN. Si X es variable aleatoria, entonces $|X|$ es variable aleatoria.

Demostración. Si $x \geq 0$, entonces $(|X| \leq x) = (-x \leq X \leq x)$, y si $x < 0$, entonces $(|X| \leq x) = \emptyset \in \mathcal{F}$, de modo que $|X|$ es variable aleatoria. Alternativamente se puede escribir $|X| = X^+ + X^-$, y por lo expuesto anteriormente obtener la misma conclusión. □

Se muestra a continuación que en general el recíproco de la proposición anterior es falso, esto es, si $X : \Omega \rightarrow \mathbb{R}$ es una función tal que $|X|$ es variable aleatoria, entonces no necesariamente X es variable aleatoria.

EJEMPLO. Considere el espacio muestral $\Omega = \{-1, 0, 1\}$ junto con la σ -álgebra $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. Sea $X : \Omega \rightarrow \mathbb{R}$ la función identidad $X(\omega) = \omega$. Entonces $|X|$ es variable aleatoria pues para cualquier conjunto Boreliano B ,

$$|X|^{-1}B = \begin{cases} \Omega & \text{si } 0, 1 \in B, \\ \{-1, 1\} & \text{si } 0 \notin B \text{ y } 1 \in B, \\ \{0\} & \text{si } 0 \in B \text{ y } 1 \notin B, \\ \emptyset & \text{si } 0, 1 \notin B. \end{cases}$$

Es decir, $|X|^{-1}B$ es un elemento de \mathcal{F} . Sin embargo X no es variable aleatoria pues el conjunto $X^{-1}\{-1\} = \{-1\}$ no es un elemento de \mathcal{F} . \square

Ahora consideraremos algunas operaciones límite en sucesiones infinitas de variables aleatorias. Sólo consideraremos variables aleatorias con valores finitos, de modo que impondremos condiciones sobre la finitud del resultado al tomar tales operaciones límite.

PROPOSICIÓN. Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias tales que para cada ω en Ω , los números

$$\sup \{X_1(\omega), X_2(\omega), \dots\} \quad \text{e} \quad \inf \{X_1(\omega), X_2(\omega), \dots\}$$

son finitos. Entonces las funciones $\sup_{n \geq 0} \{X_n\}$ e $\inf_{n \geq 0} \{X_n\}$ son variables aleatorias.

Demostración. Para cualquier número real x ,

$$\begin{aligned} (\sup_{n \geq 0} X_n \leq x) &= \bigcap_{n=1}^{\infty} (X_n \leq x), \\ \text{e} \quad (\inf_{n \geq 0} X_n \geq x) &= \bigcap_{n=1}^{\infty} (X_n \geq x). \end{aligned}$$

\square

El siguiente resultado hace uso de las operaciones de límite superior e inferior para sucesiones numéricas, el lector puede encontrar una revisión breve de estas operaciones al final del texto.

PROPOSICIÓN. Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias tales que para cada ω en Ω , los números

$$\limsup \{X_1(\omega), X_2(\omega), \dots\} \quad \text{y} \quad \liminf \{X_1(\omega), X_2(\omega), \dots\}$$

son finitos. Entonces las funciones $\limsup_{n \rightarrow \infty} X_n$ y $\liminf_{n \rightarrow \infty} X_n$ son variables aleatorias.

Demostración. Esto es consecuencia de la proposición anterior pues

$$\begin{aligned} \limsup_{n \rightarrow \infty} X_n &= \inf_k (\sup_{n \geq k} X_n), \\ \text{y} \quad \liminf_{n \rightarrow \infty} X_n &= \sup_k (\inf_{n \geq k} X_n). \end{aligned}$$

□

Finalmente demostramos que el límite de una sucesión de variables aleatorias convergente es variable aleatoria.

PROPOSICIÓN. Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias tales que $\lim_{n \rightarrow \infty} X_n(\omega)$ existe y es finito para cada $\omega \in \Omega$. Entonces la función $\lim_{n \rightarrow \infty} X_n$ es una variable aleatoria.

Demostración. Como el límite de X_n existe, los límites superior e inferior de esta sucesión coinciden. Entonces por lo demostrado antes, el límite de X_n es variable aleatoria. □

2.2. Función de distribución

Toda variable aleatoria tiene asociada una función llamada función de distribución. En esta sección se define esta importante función y se demuestran algunas de sus propiedades.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN). La función de distribución de una variable aleatoria X es la función $F(x) : \mathbb{R} \rightarrow [0, 1]$, definida como sigue

$$F(x) = P(X \leq x).$$

Cuando sea necesario especificar la variable aleatoria en cuestión se escribe $F_X(x)$, pero en general se omite el subíndice X cuando no haya posibilidad de confusión. El argumento de la función es la letra minúscula x que puede tomar cualquier valor real. Por razones obvias a esta función se le conoce también con el nombre de *función de acumulación de probabilidad*, o *función de probabilidad acumulada*. Observe que la función de distribución de una variable aleatoria está definida sobre la totalidad del conjunto de números reales, y siendo una probabilidad, toma valores en el intervalo $[0, 1]$.

La función de distribución es importante pues, como se ilustrará más adelante, contiene ella toda la información de la variable aleatoria y la correspondiente medida de probabilidad. Veremos a continuación algunas propiedades básicas de esta función, en una de las cuales aparece la expresión $F(x+)$, que significa el límite por la derecha de la función F en el punto x . Aparecerá también la expresión $F(x-)$, que significa, de manera análoga, el límite por la izquierda de la función F en el punto x .

PROPOSICIÓN. Sea $F(x)$ la función de distribución de una variable aleatoria. Entonces

1. $\lim_{x \rightarrow +\infty} F(x) = 1.$
2. $\lim_{x \rightarrow -\infty} F(x) = 0.$
3. Si $x_1 \leq x_2$, entonces $F(x_1) \leq F(x_2).$
4. $F(x)$ es continua por la derecha, es decir, $F(x+) = F(x).$

Demostración.

1. Sea x_1, x_2, \dots una sucesión cualquiera de números reales creciente a infinito, y sean los eventos $A_n = (X \leq x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos creciente cuyo límite es Ω . Por la propiedad de continuidad

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\Omega) = 1.$$

Dado que \mathbb{R} es un espacio métrico, lo anterior implica que $F(x)$ converge a uno cuando x tiende a infinito.

2. Sea ahora $\{x_n : n \in \mathbb{N}\}$ una sucesión cualquiera de números reales decreciente a menos infinito, y sean los eventos $A_n = (X \leq x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos decreciente al conjunto vacío. Nuevamente por la propiedad de continuidad

$$\lim_{n \rightarrow \infty} F(x_n) = \lim_{n \rightarrow \infty} P(A_n) = P(\emptyset) = 0.$$

Por lo tanto, $F(x)$ converge a cero cuando x tiende a menos infinito.

3. Para $x_1 \leq x_2$,

$$\begin{aligned} F(x_1) &\leq F(x_1) + P(x_1 < X \leq x_2) \\ &= P[(X \leq x_1) \cup (x_1 < X \leq x_2)] \\ &= P(X \leq x_2) \\ &= F(x_2). \end{aligned}$$

4. Sea x_1, x_2, \dots una sucesión cualquiera de números reales no negativos y decreciente a cero. Entonces

$$F(x + x_n) = F(x) + P(x < X \leq x + x_n),$$

en donde $A_n = (x < X \leq x + x_n)$ es una sucesión de eventos decreciente al conjunto vacío. Por lo tanto $\lim_{n \rightarrow \infty} F(x + x_n) = F(x)$. Es decir $F(x+) = F(x)$.

□

Se tiene además la siguiente definición general de función de distribución, no haciendo referencia a variables aleatorias ni a espacios de probabilidad particulares.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN). Una función $F(x) : \mathbb{R} \rightarrow [0, 1]$ es llamada función de distribución si cumple las cuatro propiedades anteriores.

Una especie de recíproco de la última proposición es válido y ello justifica la importancia de la función de distribución. Se enuncia a continuación este interesante resultado cuya demostración omitiremos y puede encontrarse por ejemplo en [15].

PROPOSICIÓN. Sea $F(x) : \mathbb{R} \rightarrow [0, 1]$ una función de distribución. Entonces existe un espacio de probabilidad (Ω, \mathcal{F}, P) y una variable aleatoria X cuya función de distribución es $F(x)$.

Por lo tanto basta dar una función de distribución específica para saber que existe un cierto espacio de probabilidad y una variable aleatoria definida sobre él y cuya función de distribución es la especificada. Este es el punto de vista que a menudo se adopta en el estudio de las variables aleatorias, quedando un espacio de probabilidad no especificado en el fondo como elemento base en todas las consideraciones.

A continuación se presentan algunos ejemplos gráficos de funciones de distribución. La gráfica de la izquierda en la Figura 2.3 corresponde a la función de distribución de una variable aleatoria discreta, y la gráfica de la derecha muestra el comportamiento típico de una función de distribución continua. También pueden presentarse situaciones como la que se muestra en la Figura 2.4, y que corresponde al caso de una variable aleatoria mixta. La definición de variable aleatoria discreta, continua y mixta aparece en la siguiente sección.

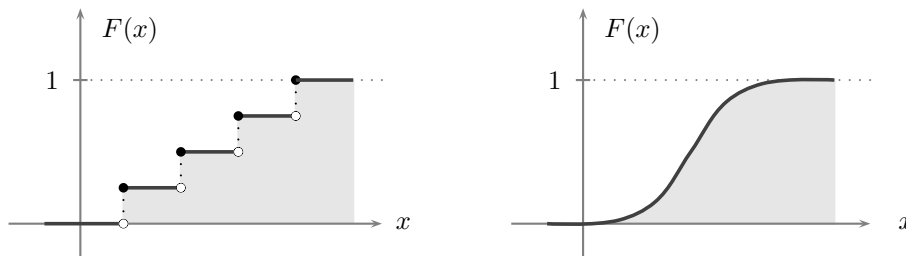


Figura 2.3: Funciones de distribución discreta y continua.

Se demuestran ahora algunas otras propiedades que establecen la forma de calcular probabilidades usando la función de distribución.

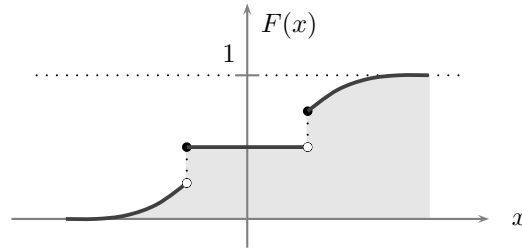


Figura 2.4: Una función de distribución mixta.

PROPOSICIÓN. Sea X una variable aleatoria con función de distribución $F(x)$. Para cualesquiera números reales $a < b$,

1. $P(X < a) = F(a-)$.
2. $P(X = a) = F(a) - F(a-)$.
3. $P(a < X \leq b) = F(b) - F(a)$.
4. $P(a \leq X \leq b) = F(b) - F(a-)$.
5. $P(a < X < b) = F(b-) - F(a)$.
6. $P(a \leq X < b) = F(b-) - F(a-)$.

Demostración.

1. Sea x_1, x_2, \dots una sucesión de números reales positivos y decreciente a cero. Sea A_n el evento $(X \leq a - x_n)$. Entonces $\{A_n : n \in \mathbb{N}\}$ es una sucesión de eventos decreciente al evento $(X < a)$. Por la propiedad de continuidad

$$P(X < a) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(a - x_n) = F(a-).$$

2. Simplemente se escribe

$$P(X = a) = P(X \leq a) - P(X < a) = F(a) - F(a-).$$

3.–6. Estas igualdades se sigue directamente de las dos primeras.

□

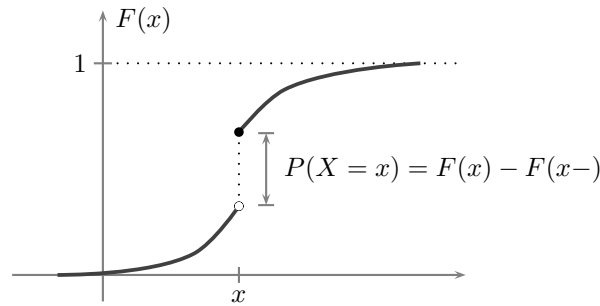


Figura 2.5: Una discontinuidad de $F(x)$ y su significado.

Observe que como $F(x)$ es una función no decreciente y continua por la derecha, la probabilidad $P(X = x)$ es igual a $F(x) - F(x-)$, que representa el tamaño del salto o discontinuidad de la función de distribución en el punto x . Esto se muestra en la Figura 2.5. En consecuencia, cuando $F(x)$ es una función continua y para $a < b$,

$$\begin{aligned} F(b) - F(a) &= P(a < X \leq b) \\ &= P(a \leq X \leq b) \\ &= P(a < X < b) \\ &= P(a \leq X < b). \end{aligned}$$

Es decir, cuando $F(x)$ es una función continua, incluir o excluir los extremos de un intervalo no afecta el valor de la probabilidad de dicho intervalo. Por

lo tanto, para cualquier número real x , la probabilidad del evento $(X = x)$ es cero. Finalizamos esta sección con un resultado interesante cuya prueba es sorprendentemente simple.

PROPOSICIÓN. Toda función de distribución tiene a lo sumo un número numerable de discontinuidades.

Demostración. Sea D el conjunto de puntos de discontinuidad de una función de distribución $F(x)$. Para cada número natural n defina los subconjuntos

$$D_n = \left\{ x \in D : \frac{1}{n+1} < F(x) - F(x-) \leq \frac{1}{n} \right\}.$$

Cada conjunto D_n tiene a lo sumo n elementos. Como $D = \bigcup_{n=1}^{\infty} D_n$, se concluye que D es numerable. \square

2.3. Tipos de variables aleatorias

Las variables aleatorias se clasifican en varios tipos dependiendo de las características de la correspondiente función de distribución. Al menos existen tres tipos: discretas, continuas, y mezclas de las dos anteriores. Veamos su definición.

DEFINICIÓN. (VARIABLE ALEATORIA DISCRETA). La variable aleatoria X se llama discreta si su correspondiente función de distribución $F(x)$ es una función constante por pedazos. Sean x_1, x_2, \dots los puntos de discontinuidad de $F(x)$. En cada uno de estos puntos el tamaño de la discontinuidad es $P(X = x_i) = F(x_i) - F(x_i-) > 0$. A la función $f(x)$ que indica estos incrementos se le llama función de probabilidad de X , y se define como sigue

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_1, x_2, \dots \\ 0 & \text{otro caso.} \end{cases} \quad (2.2)$$

La función de distribución se reconstruye de la forma siguiente

$$F(x) = \sum_{u \leq x} f(u).$$

En este caso se dice también que la función de distribución es *discreta*, además la función de probabilidad $f(x)$ siempre existe, y se le llama también *función de masa de probabilidad*. También se acostumbra usar el término *función de densidad*, como una analogía con el caso de variables aleatorias continuas definidas más adelante. Cuando sea necesario especificarlo se escribe $f_X(x)$ en lugar de $f(x)$. Observe que la función de probabilidad $f(x)$ es una función no negativa que suma uno en el sentido $\sum_i f(x_i) = 1$. Recíprocamente, toda función de la forma (2.2) que cumpla estas dos propiedades se le llama *función de probabilidad*, sin que haya necesariamente una variable aleatoria de por medio. Veamos ahora el caso continuo.

DEFINICIÓN. (VARIABLE ALEATORIA CONTINUA). La variable aleatoria X se llama continua si su correspondiente función de distribución es una función continua.

En tal caso también se dice que la distribución es *continua*. Las distribuciones continuas se clasifican a su vez en distribuciones *absolutamente continuas*

y distribuciones *singulares*.

DEFINICIÓN. (VARIABLE ALEATORIA ABSOLUTAMENTE CONTINUA). La variable aleatoria continua X con función de distribución $F(x)$ se llama absolutamente continua, si existe una función no negativa e integrable f tal que para cualquier valor de x se cumple

$$F(x) = \int_{-\infty}^x f(u) du. \quad (2.3)$$

En tal caso a la función $f(x)$ se le llama función de densidad de X .

Aún cuando exista una función no negativa e integrable f que cumpla (2.3), ésta puede no ser única, pues basta modificarla en un punto para que sea ligeramente distinta pero aún así seguir cumpliendo (2.3). A pesar de ello, nos referiremos a la función de densidad como si ésta fuera única, y ello se justifica por el hecho de que las probabilidades son las mismas, ya sea usando una función de densidad o modificaciones de ella que cumplan (2.3). Es claro que la función de densidad de una variable aleatoria absolutamente continua es no negativa y su integral sobre toda la recta real es uno. Recíprocamente, toda función $f(x)$ no negativa que integre uno en \mathbb{R} se llama *función de densidad*. Si X es absolutamente continua con función de distribución $F(x)$ y función de densidad continua $f(x)$, entonces el teorema fundamental del cálculo establece que, a partir de (2.3), $F'(x) = f(x)$. Además, la probabilidad de que X tome un valor en el intervalo (a, b) es el área bajo la función de densidad sobre dicho intervalo. Esto se ilustra en la Figura 2.6, la probabilidad es la misma si se incluyen o excluyen los extremos del intervalo.

Pueden construirse ejemplos de variables aleatorias continuas que no tienen función de densidad, es decir, que no existe una función f no negativa e integrable que cumpla (2.3) para cualquier número real x . En tales situaciones se dice que la distribución es *singular*.

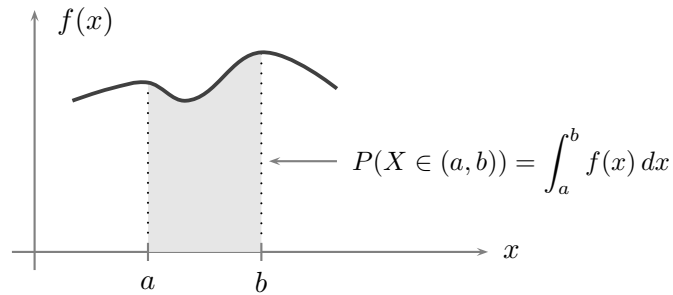


Figura 2.6: La probabilidad como el área bajo la función de densidad.

DEFINICIÓN. (VARIABLE ALEATORIA SINGULAR). La variable aleatoria continua X , o su correspondiente función de distribución $F(x)$, se llama singular si $F'(x) = 0$ casi seguramente.

El término “casi seguramente” que aparece en esta definición se refiere a que la igualdad se cumple en todos los puntos x excepto en un conjunto cuya medida de Lebesgue es cero. Las distribuciones singulares son un poco más delicadas de estudiar y no haremos mayor énfasis en ellas. La *distribución de Cantor* es un ejemplo de este tipo de distribuciones y se construye mediante un proceso límite. Los detalles pueden encontrarse en [13] o [19].

DEFINICIÓN. (VARIABLE ALEATORIA MIXTA). Una variable aleatoria que no es discreta ni continua se llama variable aleatoria mixta.

No es difícil encontrar situaciones en donde la variable aleatoria en estudio es mixta, el siguiente ejemplo es una muestra de ello.

EJEMPLO (UNA VARIABLE ALEATORIA QUE NO ES DISCRETA NI CONTINUA). Sea X una variable aleatoria con función de distribución

$$F_X(x) = \begin{cases} 1 - e^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Como la función $F_X(x)$ es continua, entonces la variable aleatoria X es continua. Sea $M > 0$ una constante. Las gráficas de las funciones de distribución de las variables X y la constante M (vista como variable aleatoria), se muestran en la Figura 2.7.

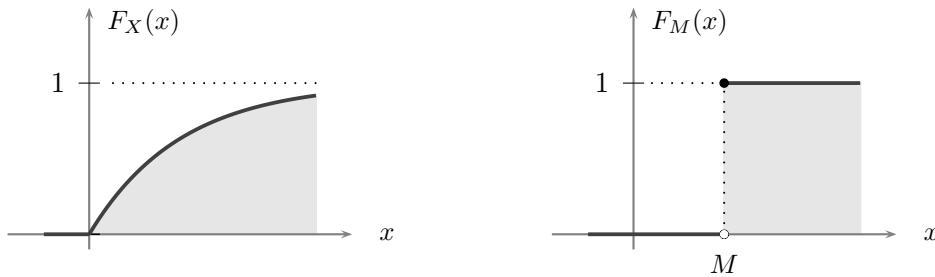


Figura 2.7: Funciones de distribución de la variable X y la constante M .

Sea $Y = \min\{X, M\}$. Puede comprobarse que la función de distribución de Y es

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ 1 - e^{-y} & \text{si } 0 < y < M, \\ 1 & \text{si } y \geq M, \end{cases}$$

con gráfica como en la Figura 2.8. Es claro que esta función de distribución no es constante por pedazos pues es creciente en el intervalo $(0, M)$, por lo tanto no es discreta, y tampoco es continua pues tiene una discontinuidad en $y = M$. Por lo tanto Y es una variable aleatoria que no es discreta ni continua.

▪

Finalmente enunciamos un resultado general cuya demostración puede encontrarse en [7] o [13].

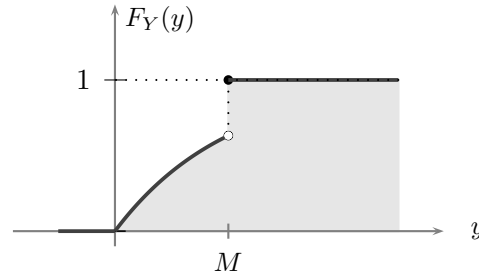


Figura 2.8: Función de distribución de la variable $Y = \min\{X, M\}$.

PROPOSICIÓN. Toda función de distribución $F(x)$ se puede escribir como una combinación lineal convexa de una función de distribución discreta $F^d(x)$ y otra continua $F^c(x)$, es decir, admite la siguiente representación

$$F(x) = \alpha F^d(x) + (1 - \alpha)F^c(x),$$

en donde $0 \leq \alpha \leq 1$.

En todos los casos que consideraremos en este texto la distribución continua de esta descomposición será absolutamente continua. En el caso general, esta distribución continua puede a su vez escribirse como otra combinación lineal convexa entre una distribución absolutamente continua y una distribución continua singular. Esto lleva al resultado general de que cualquier distribución puede escribirse como una combinación lineal convexa de los tres tipos básicos de distribuciones.

EJEMPLO. Considere nuevamente la función de distribución de la variable $Y = \min\{X, M\}$ analizada en el ejemplo anterior. Hemos visto que esta distribución no es discreta ni continua, sin embargo puede descomponerse en la combinación lineal convexa

$$F_Y(y) = e^{-M} F^d(y) + (1 - e^{-M}) F^c(y),$$

en donde $F^d(y)$ es la distribución discreta de la variable constante M , y

$F^c(y)$ es la distribución continua

$$F_Y^c(y) = \begin{cases} 0 & \text{si } y \leq 0, \\ \frac{1 - e^{-y}}{1 - e^{-M}} & \text{si } 0 < y < M, \\ 1 & \text{si } y \geq M. \end{cases}$$

Igualdad de variables aleatorias

Dos variables aleatorias X y Y son estrictamente iguales si para cada ω se cumple $X(\omega) = Y(\omega)$. Existen, sin embargo, otras formas más débiles de igualdad que enunciaremos a continuación.

DEFINICIÓN. (IGUALDAD DE VARIABLES ALEATORIAS). Se dice que dos variables aleatorias X y Y son

- a) iguales casi seguramente, y se escribe $X = Y$ c.s., o bien $X \stackrel{c.s.}{=} Y$, si se cumple que $P(X = Y) = 1$. Más generalmente, un evento ocurre casi seguramente si su probabilidad es uno.
- b) iguales en distribución, y se escribe $X \stackrel{d}{=} Y$, si sus correspondientes funciones de distribución coinciden, es decir, si $F_X(x) = F_Y(x)$ para cada número real x .

Es interesante observar que la igualdad casi segura es más fuerte que la igualdad en distribución, es decir, si X y Y son iguales casi seguramente, entonces son iguales en distribución. Sin embargo, si X y Y tienen la misma distribución, entonces no necesariamente son iguales casi seguramente. A menos que se indique lo contrario, cuando aparezca una expresión de igualdad entre variables aleatorias, se considera que la igualdad es válida en el sentido fuerte, es decir, casi seguro.

EJERCICIO. Sean X y Y dos variables aleatorias. Demuestre que el conjunto $(X = Y)$ es un evento. En consecuencia tiene sentido calcular la probabilidad de tal conjunto. \cdot

EJERCICIO. Demuestre que si $X = Y$ c.s., entonces $X \stackrel{d}{=} Y$. Por el contrario, demuestre que si $X \stackrel{d}{=} Y$, entonces no necesariamente $X = Y$ c.s. Considere por ejemplo la variable X tal que $P(X = -1) = P(X = 1) = 1/2$, y defina $Y = -X$. \cdot

2.4. Integral de Riemann-Stieltjes

En esta sección se define la integral de Riemann-Stieltjes. Esta es una integral de la forma

$$\int_a^b h(x) dF(x),$$

en donde las funciones $h(x)$ y $F(x)$ deben cumplir ciertas condiciones para que la integral tenga sentido y esté bien definida. Esta integral es una generalización de la integral usual de Riemann. Al integrando $h(x)$ se le pide inicialmente que sea una función acotada en el intervalo $(a, b]$, aunque después se omitirá esta condición. A la función integradora $F(x)$ se le pide que sea continua por la derecha, monótona no decreciente y tal que $F(\infty) - F(-\infty) < M$, para algún número $M > 0$. Observe que $F(x)$ debe cumplir propiedades semejantes a las de una función de distribución, y de hecho la notación es la misma. Esto no es coincidencia pues usaremos las funciones de distribución como funciones integradoras.

Presentamos a continuación la definición de la integral de Riemann-Stieltjes bajo las condiciones arriba señaladas. En [15] puede encontrarse una exposición más completa y rigurosa de esta integral. Sea $\{a = x_0 < x_1 < \cdots < x_n = b\}$ una partición finita del intervalo $(a, b]$, y defina

$$\begin{aligned} \bar{h}(x_i) &= \sup \{h(x) : x_{i-1} < x \leq x_i\}, \\ \text{y } \underline{h}(x_i) &= \inf \{h(x) : x_{i-1} < x \leq x_i\}. \end{aligned}$$

Se define la *suma superior e inferior* de Riemann-Stieltjes como sigue

$$\begin{aligned}\overline{S}_n &= \sum_{i=1}^n \bar{h}(x_i) [F(x_i) - F(x_{i-1})], \\ \text{y } \underline{S}_n &= \sum_{i=1}^n \underline{h}(x_i) [F(x_i) - F(x_{i-1})].\end{aligned}$$

Ahora se toma el límite cuando n tiende a infinito de tal forma que la longitud máx $\{|x_i - x_{i-1}| : 1 \leq i \leq n\}$ tienda a cero. Si sucede que

$$-\infty < \lim_{n \rightarrow \infty} \underline{S}_n = \lim_{n \rightarrow \infty} \overline{S}_n < \infty,$$

entonces a este valor común se le llama la *integral de Riemann-Stieltjes* de la función $h(x)$ respecto de la función $F(x)$ sobre el intervalo $(a, b]$, y se le denota por

$$\int_a^b h(x) dF(x),$$

Cuando la función $h(x)$ no es acotada se hace uso de la función auxiliar

$$h_N(x) = \begin{cases} -N & \text{si } h(x) < -N, \\ h(x) & \text{si } |h(x)| \leq N, \\ N & \text{si } h(x) > N. \end{cases}$$

Y entonces se define

$$\int_a^b h(x) dF(x) = \lim_{N \rightarrow \infty} \int_a^b h_N(x) dF(x),$$

cuando este límite existe. Se puede extender la definición de esta integral de la siguiente forma

$$\int_{-\infty}^{\infty} h(x) dF(x) = \lim_{a, b \rightarrow \infty} \int_a^b h(x) dF(x),$$

cuando el límite del lado derecho exista y esté bien definido.

La integral de Riemann-Stieltjes tiene varias propiedades semejantes a la integral de Riemann, enunciaremos a continuación algunas de ellas. Primeramente es lineal tanto en el integrando como en el integrador, es decir, si α es constante, entonces

$$\begin{aligned} \text{a) } & \int_a^b (\alpha h_1(x) + h_2(x)) dF(x) = \alpha \int_a^b h_1(x) dF(x) + \int_a^b h_2(x) dF(x). \\ \text{b) } & \int_a^b h(x) d(\alpha F_1(x) + F_2(x)) = \alpha \int_a^b h(x) dF_1(x) + \int_a^b h(x) dF_2(x). \end{aligned}$$

Cuando $h(x)$ tiene primera derivada continua se cumple la fórmula

$$\text{c) } \int_a^b h(x) dF(x) = h(b)F(b) - h(a)F(a) - \int_a^b F(x)h'(x) dx.$$

De particular importancia en la teoría de la probabilidad son los siguientes dos casos particulares. Cuando $F(x)$ es diferenciable se tiene la igualdad

$$\text{d) } \int_a^b h(x) dF(x) = \int_a^b h(x)F'(x) dx.$$

Es decir, integrar respecto de una función de distribución absolutamente continua se reduce a efectuar una integral de Riemann. El otro caso interesante ocurre cuando $h(x)$ es continua y $F(x)$ es constante excepto en los puntos x_1, x_2, \dots , en donde la función tiene saltos positivos de tamaño $p(x_1), p(x_2), \dots$ respectivamente. En este caso y suponiendo convergencia,

$$\text{e) } \int_a^b h(x) dF(x) = \sum_{i=1}^{\infty} h(x_i) p(x_i).$$

Esto significa que integrar respecto de la función de distribución de una variable aleatoria discreta se reduce a efectuar una suma. Finalmente enunciamos la propiedad que ilustra el hecho de que la integral de Riemann es

un caso particular de la integral de Riemann-Stieltjes. Cuando $F(x) = x$ se cumple

$$f) \int_a^b h(x) dF(x) = \int_a^b h(x) dx.$$

En la siguiente sección usaremos las funciones de distribución como funciones integradoras. Como toda función de distribución $F(x)$ se puede descomponer en una suma convexa $\alpha F^d(x) + (1 - \alpha)F^c(x)$, en donde $F^d(x)$ es discreta y $F^c(x)$ es continua, entonces

$$\int_a^b h(x) dF(x) = \alpha \int_a^b h(x) dF^d(x) + (1 - \alpha) \int_a^b h(x) dF^c(x).$$

En algunos casos usaremos también la integral de Riemann-Stieltjes en varias dimensiones. Por ejemplo, sean $h(x, y)$ y $F(x, y)$ funciones de dos variables, sea $\{a = x_0 < x_1 < \dots < x_n = b\}$ una partición de $(a, b]$ y sea $\{c = y_0 < y_1 < \dots < y_m = d\}$ una partición de $(c, d]$, entonces se define

$$\int_a^b \int_c^d h(x, y) dF(x, y) = \lim_{n, m} \sum_{i=1}^n \sum_{j=1}^m h(x_i, y_j) \Delta F(x_i, y_j),$$

en donde $\Delta F(x_i, y_j)$ es el “incremento” de F en el rectángulo $(x_{i-1}, x_i] \times (y_{j-1}, y_j]$. Por ahora no es clara la forma de definir este incremento pero retomaremos este concepto una vez que se haya definido a la función de distribución en dimensiones mayores.

2.5. Características numéricas

Se estudian a continuación algunas características numéricas asociadas a variables aleatorias. En particular, se definen los conceptos de esperanza, varianza y más generalmente los momentos de una variable aleatoria. Para ello haremos uso de la integral de Riemann-Stieltjes mencionada antes.

Esperanza

La esperanza de una variable aleatoria es un número que representa el promedio ponderado de sus posibles valores, se calcula como se indica a continuación.

DEFINICIÓN. (ESPERANZA). Sea X con función de distribución $F(x)$. La esperanza de X , denotada por $E(X)$, se define como el número

$$E(X) = \int_{-\infty}^{\infty} x dF(x),$$

cuando esta integral sea absolutamente convergente, es decir, cuando $\int_{-\infty}^{\infty} |x| dF(x) < \infty$, y en tal caso se dice que X es integrable, o que tiene esperanza finita.

A la esperanza se le conoce también con el nombre de *media*, *valor esperado*, *valor promedio* o *valor medio*, y en general se usa la letra griega μ (mu) para denotarla. En la teoría de la medida [5] [14] [29] se define la esperanza de una variable aleatoria o función medible X mediante una integral más general llamada *integral de Lebesgue*, y se denota por

$$\int_{\Omega} X(\omega) dP(\omega).$$

En algunas ocasiones usaremos esta expresión para tener compatibilidad en notación con la teoría general.

Cuando X es discreta con función de probabilidad $f(x)$, su esperanza, si existe, se calcula como sigue $E(X) = \sum_x xf(x)$. Si X es absolutamente continua con función de densidad $f(x)$, entonces su esperanza, si existe, es $E(X) = \int_{-\infty}^{\infty} xf(x) dx$.

EJEMPLOS.

- a) Sea X con valores en el conjunto $\{1, 2, \dots\}$, y con función de probabilidad $f(x) = P(X = x) = 1/2^x$, para $x \geq 1$. Entonces $E(X) =$

$$\sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} x/2^x = 2.$$

- b) Sea X continua con función de densidad $f(x) = 2x$, para $0 < x < 1$.
Entonces $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x 2x dx = 2/3$.

▪

La integral o suma arriba mencionados pueden no existir y en ese caso se dice que la variable aleatoria no tiene esperanza finita. El siguiente ejercicio contiene un par de ejemplos que ilustran esta situación. Véase también el ejercicio 151.

EJERCICIO. Demuestre que no existe la esperanza de X cuando su función de probabilidad o de densidad es

a) $f(x) = \frac{1}{x(x+1)}$, para $x = 1, 2, \dots$

b) $f(x) = 1/x^2$, para $x > 1$.

▪

EJEMPLO. Sea X una variable aleatoria con la siguiente función de distribución. La forma de esta función puede apreciarse más fácilmente a través de su gráfica, la cual se muestra en la Figura 2.9.

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x/4 & \text{si } 0 \leq x < 1, \\ 2/4 & \text{si } 1 \leq x < 2, \\ 1/4 + x/4 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3. \end{cases}$$

De acuerdo a las propiedades de la integral de Riemann-Stieltjes, la espe-

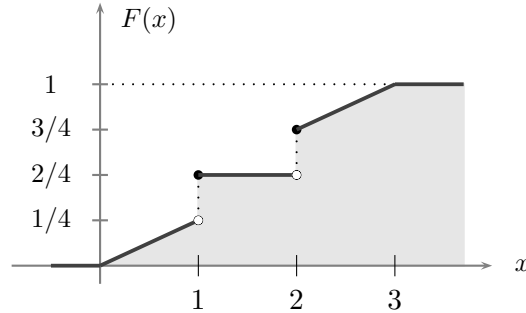


Figura 2.9: Una función de distribución mixta.

ranza de X es entonces

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x dF(x) \\ &= \int_0^1 x \frac{1}{4} dx + 1 \left(\frac{2}{4} - \frac{1}{4} \right) + 2 \left(\frac{3}{4} - \frac{2}{4} \right) + \int_2^3 x \frac{1}{4} dx. \end{aligned}$$

Después de algunos cálculos se encuentra que la esperanza es $15/4$. Observe la forma mixta en la que esta integral es calculada: en las partes crecientes se calcula como si fuera una distribución continua, después se añaden los puntos de discontinuidad ponderados por el tamaño del salto. \blacksquare

Con frecuencia surge el problema de calcular esperanzas de funciones de variables aleatorias, es decir, si X es una variable aleatoria y $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función Borel medible, entonces $g(X)$ es una variable aleatoria y el problema es encontrar su esperanza. Usando directamente la definición, la esperanza de $g(X)$ se calcula del siguiente modo:

$$E[g(X)] = \int_{-\infty}^{\infty} x dF_{g(X)}(x),$$

pero ello requiere encontrar primero la distribución de $g(X)$, lo cual puede no ser fácil en muchos casos. Afortunadamente se cuenta con el siguiente resultado que establece una forma muy conveniente de calcular la esperanza de

$g(X)$, sin conocer su distribución, pero suponiendo conocida la distribución de X .

TEOREMA. (ESPERANZA DE UNA FUNCIÓN DE UNA V.A.) Sea X con función de distribución $F_X(x)$, y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función Borel medible tal que $g(X)$ tiene esperanza finita. Entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

La demostración de este resultado en general no es sencilla y la omitiremos, aunque un camino cómodo que puede adoptarse es aceptar la fórmula anterior como la definición de la esperanza de $g(X)$. En particular, cuando la función g es la identidad, se recupera la definición básica de esperanza. Por otro lado, cuando X es discreta, la demostración del teorema resulta no ser complicada.

EJERCICIO. Sea X una variable aleatoria discreta con valores en el conjunto $\{x_1, x_2, \dots\}$, y sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función Borel medible tal que $g(X)$ tiene esperanza finita. Demuestre que

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)P(X = x_i).$$

▪

Se establecen a continuación algunas propiedades de la esperanza.

PROPOSICIÓN. (PROPIEDADES DE LA ESPERANZA). Sean X y Y con esperanza finita, y sea c una constante. Entonces

1. $E(c) = c$.
2. $E(cX) = cE(X)$.
3. Si $X \geq 0$, entonces $E(X) \geq 0$.
4. Si $X \leq Y$, entonces $E(X) \leq E(Y)$.
5. $E(X + Y) = E(X) + E(Y)$.

Las demostraciones de las primeras cuatro propiedades son sencillas pues se siguen directamente de la definición. La última propiedad es fácilmente demostrable en el caso discreto. El caso general será demostrado más adelante.

EJERCICIO. Sean X y Y discretas ambas con esperanza finita. Demuestre directamente que $E(X + Y) = E(X) + E(Y)$. ■

PROPOSICIÓN. Sea X con función de distribución $F(x)$, la cual admite la descomposición

$$F(x) = \alpha F^d(x) + (1 - \alpha)F^c(x),$$

en donde $\alpha \in [0, 1]$, $F^d(x)$ es una función de distribución discreta, y $F^c(x)$ es una función de distribución continua. Sea X_d con distribución $F^d(x)$, y sea X_c con distribución $F^c(x)$. Entonces X tiene esperanza finita si, y sólo si, tanto X_d como X_c tienen esperanza finita, y en tal caso,

$$E(X) = \alpha E(X_d) + (1 - \alpha)E(X_c).$$

Este resultado es inmediato de demostrar usando la propiedad de linealidad de la integral de Riemann-Stieltjes respecto de la función integradora.

Varianza

La varianza de una variable aleatoria es una medida del grado de dispersión de los diferentes valores tomados por la variable, su definición es la siguiente.

DEFINICIÓN. (VARIANZA). La varianza de una variable aleatoria X , denotada por $\text{Var}(X)$, se define como la siguiente esperanza, si ésta existe,

$$\text{Var}(X) = E(X - E(X))^2.$$

Cuando X es discreta con función de probabilidad $f(x)$ y esperanza finita μ , la varianza de X , cuando existe, se calcula como sigue $\text{Var}(X) = \sum_x (x - \mu)^2 f(x)$. Si X es absolutamente continua con función de densidad $f(x)$ y esperanza finita μ , entonces la varianza de X , cuando existe, es $\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$. La varianza se denota regularmente por el símbolo σ^2 (sigma cuadrada). A la raíz cuadrada positiva de $\text{Var}(X)$ se le llama *desviación estándar*, y se le denota naturalmente por σ . Nuevamente hay casos en los que la varianza no es finita, y en esas situaciones se dice que la variable aleatoria no tiene varianza. Observe que para calcular la varianza se necesita conocer primero la esperanza.

EJERCICIO. Demuestre que la varianza de una variable aleatoria con la siguiente función de densidad no existe.

$$f(x) = \begin{cases} 2/x^3 & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

Enunciamos a continuación algunas propiedades de la varianza.

PROPOSICIÓN. (PROPIEDADES DE LA VARIANZA). Sean X y Y con varianza finita, y sea c una constante. Entonces

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(c) = 0$.
3. $\text{Var}(cX) = c^2 \text{Var}(X)$.
4. $\text{Var}(X + c) = \text{Var}(X)$.
5. $\text{Var}(X) = E(X^2) - E^2(X)$.
6. En general, $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$.

La demostración de estas propiedades es sencilla pues todas ellas, excepto la última, se siguen directamente de la definición y de la propiedad lineal de la esperanza. Para la última propiedad puede tomarse $Y = X$, con $\text{Var}(X) \neq 0$, y verificarse la no igualdad. Otras propiedades de la varianza aparecen más adelante.

EJERCICIO. Demuestre que $\text{Var}(X) = E(X(X - 1)) - E(X)(E(X) - 1)$.

Momentos

Los momentos de una variable aleatoria son números que representan algunas características de la distribución de probabilidad asociada. Bajo ciertas condiciones el conjunto de momentos determinan de manera única a la distribución de probabilidad.

DEFINICIÓN. (MOMENTOS). Sea X una variable aleatoria con esperanza μ y sea n un número natural. Cuando existe, el número

1. $E(X^n)$ es el n -ésimo momento de X .
2. $E|X|^n$ es el n -ésimo momento absoluto de X .
3. $E[(X - \mu)^n]$ es el n -ésimo momento central de X .
4. $E|X - \mu|^n$ es el n -ésimo momento central absoluto de X .
5. $E[X(X - 1) \cdots (X - n + 1)]$ es el n -ésimo momento factorial de X .

Observe que el primer momento es la esperanza, y el segundo momento central es la varianza. En algunos textos al n -ésimo momento se le denota por μ'_n , mientras que el n -ésimo momento central es μ_n . En el capítulo sobre funciones generadoras se estudian ciertas funciones asociadas a las distribuciones de probabilidad, y a través de las cuales los momentos de una variable aleatoria pueden ser encontrados, cuando existen, de manera más eficiente.

El *problema de los momentos* consiste en determinar condiciones necesarias y suficientes para que los momentos de una variable aleatoria determinen de manera única su distribución de probabilidad. Por ejemplo, puede demostrarse que si X es tal que los números $E(X), E(X^2), \dots$ son todos finitos y si se cumple que la serie

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$$

es absolutamente convergente para algún $t > 0$, entonces la sucesión de momentos determina de manera única a la distribución de X . Las condiciones mencionadas son suficientes pero no necesarias.

Cuantiles

DEFINICIÓN. (CUANTIL). Sea p un número real cualquiera en el intervalo unitario $(0, 1)$. Se le llama cuantil de orden p de una variable aleatoria X o de su distribución, a cualquier número x_p que cumpla las condiciones

$$\begin{aligned} P(X \leq x_p) &\geq p, \\ \text{y } P(X \geq x_p) &\geq 1 - p. \end{aligned}$$

Es decir, el cuantil de orden p es aquel número que acumula a su izquierda una probabilidad mayor o igual a p , y al mismo tiempo acumula a su derecha una probabilidad de por lo menos $1 - p$. En general este número no es necesariamente único. Sin embargo, cuando la correspondiente función de distribución es estrictamente creciente, se cumple que el cuantil de cualquier orden es único.

A los cuantiles de orden $1/4, 1/2$ y $3/4$ se les llama también *cuartiles*. En particular al cuantil de orden $1/2$ se le llama *mediana*. Es decir, la mediana es aquel número m que cumple las desigualdades

$$\begin{aligned} P(X \leq m) &\geq 1/2, \\ \text{y } P(X \geq m) &\geq 1/2. \end{aligned}$$

La mediana de una variable aleatoria es una medida de tendencia central que permite dividir en dos partes iguales a la distribución de probabilidad cuando ésta es continua y estrictamente creciente. Usando el concepto de mediana ejemplificaremos la posible no unicidad de los cuantiles.

EJEMPLO. Sea X es una variable aleatoria discreta tal que $P(X = 1) = 1/2$, y $P(X = 0) = 1/2$. Cualquier número en el intervalo $[0, 1]$ es una mediana de X . ▪

Moda

La moda es otra característica numérica de las variables aleatorias, y se define únicamente para distribuciones discretas o absolutamente continuas de la siguiente forma.

DEFINICIÓN. (MODA). La moda de una variable aleatoria o de su distribución, discreta o absolutamente continua, es aquel punto donde la función de densidad tiene un máximo local.

Por ejemplo, si X es una variable aleatoria discreta con valores $x_1 < x_2 < x_3 < \dots$, y con probabilidades respectivas p_1, p_2, p_3, \dots , entonces X tiene una moda en el punto x_k si $p_{k-1} \leq p_k \geq p_{k+1}$. Es evidente que pueden existir varias modas para una misma variable aleatoria. Cuando la moda es única se dice que la distribución es *unimodal*, y cuando hay varias modas se dice que es *multimodal*.

2.6. Distribuciones discretas

En esta sección se estudian algunas distribuciones discretas de probabilidad de uso común. Estas distribuciones son ejemplos particulares de medidas de probabilidad concentradas en un conjunto discreto de números reales. Se presentan estos ejemplos sin hacer mayor énfasis en las aplicaciones de los modelos. En el Apéndice A, al final del libro, aparecen algunas otras distribuciones de probabilidad.

DISTRIBUCIÓN UNIFORME DISCRETA. La variable X tiene una distribución uniforme sobre el conjunto $\{x_1, \dots, x_n\}$ si la probabilidad de que X tome cualquiera de estos valores es $1/n$. Esta distribución surge en espacios de probabilidad *equiprobables*, esto es, en situaciones en donde se tienen n resultados diferentes y todos ellos tienen la misma probabilidad de ocurrir. Los juegos de lotería justos son un ejemplo donde puede aplicarse esta dis-

tribución. Se escribe $X \sim \text{unif}\{x_1, \dots, x_n\}$, y su función de probabilidad es

$$f(x) = \begin{cases} 1/n & \text{si } x = x_1, \dots, x_n, \\ 0 & \text{otro caso.} \end{cases}$$

Por ejemplo, la función de probabilidad uniforme sobre el conjunto $\{1, \dots, 5\}$ tiene gráfica como en la Figura 2.10. Es fácil ver que, en el caso general,

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i,$$

y

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - E(X))^2.$$

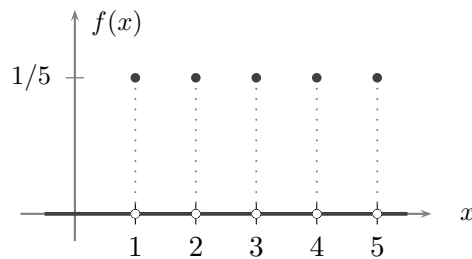


Figura 2.10: Función de probabilidad $\text{unif}\{1, \dots, 5\}$.

DISTRIBUCIÓN BERNOULLI. Un *ensayo Bernoulli* es un experimento aleatorio con únicamente dos posibles resultados, llamados genéricamente *éxito* y *fracaso*, y con probabilidades respectivas p y $1 - p$. Se define la variable aleatoria X como aquella función que lleva el resultado éxito al número 1, y el resultado fracaso al número 0. Entonces se dice que X tiene una distribución Bernoulli con parámetro $p \in (0, 1)$. Se escribe $X \sim \text{Ber}(p)$ y la correspondiente función de probabilidad es

$$f(x) = \begin{cases} 1 - p & \text{si } x = 0, \\ p & \text{si } x = 1, \\ 0 & \text{otro caso,} \end{cases}$$

cuya gráfica es como en la Figura 2.11. Es sencillo verificar que $E(X) = p$, y $\text{Var}(X) = p(1 - p)$. En particular, si A es un evento con probabilidad p , entonces la función indicadora 1_A es una variable aleatoria con distribución $\text{Ber}(p)$.

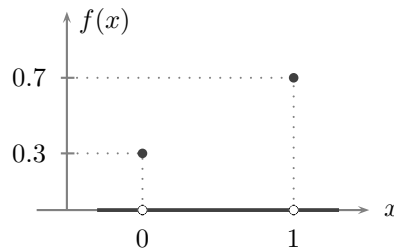


Figura 2.11: Función de probabilidad $\text{Ber}(p)$ con $p=0.7$.

DISTRIBUCIÓN BINOMIAL. Suponga que se realizan n ensayos independientes Bernoulli en donde la probabilidad de éxito en cada uno de ellos es $p \in (0, 1)$. El espacio muestral de este experimento consiste de todas las posibles sucesiones de longitud n de éxitos y fracasos. Usando el principio multiplicativo, es fácil ver que este conjunto tiene 2^n elementos. Si ahora se define la variable aleatoria X como el número de éxitos en cada una de estas sucesiones, entonces X toma los valores $0, 1, \dots, n$, y se dice que X tiene una distribución binomial con parámetros n y p . Se escribe $X \sim \text{bin}(n, p)$, y su función de probabilidad es

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}$$

Se puede demostrar que $E(X) = np$, y $\text{Var}(X) = np(1-p)$. En las gráficas de la Figura 2.12 se muestra el comportamiento de esta función de probabilidad.

DISTRIBUCIÓN GEOMÉTRICA. Suponga que se tiene una sucesión infinita de ensayos independientes Bernoulli en donde la probabilidad de éxito en

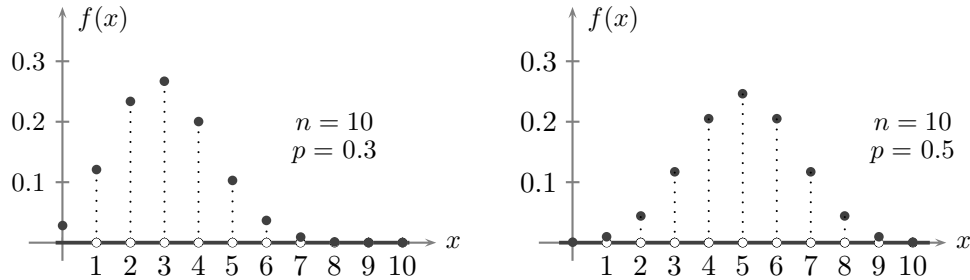


Figura 2.12: Función de probabilidad $\text{bin}(n, p)$.

cada uno de ellos es $p \in (0, 1)$. Se define X como el número de fracasos antes de obtener el primer éxito. Se dice entonces que X tiene una distribución geométrica con parámetro p . Se escribe $X \sim \text{geo}(p)$, y su función de probabilidad es

$$f(x) = \begin{cases} p(1-p)^x & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

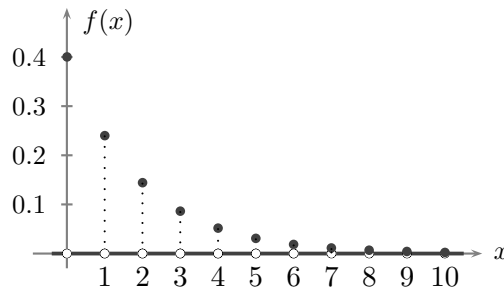


Figura 2.13: Función de probabilidad $\text{geo}(p)$ con $p=0.4$.

Para esta distribución se puede demostrar que $E(X) = (1-p)/p$, y $\text{Var}(X) = (1-p)/p^2$. En algunos textos se define también la distribución geométrica

como el número de ensayos, (y no el de fracasos), antes del primer éxito. En tal caso, la función de probabilidad es $f(x) = p(1-p)^{x-1}$, para $x = 1, 2, \dots$. La media es entonces $1/p$ y la varianza es como antes.

DISTRIBUCIÓN POISSON. La variable aleatoria discreta X tiene una distribución Poisson con parámetro $\lambda > 0$, y se escribe $X \sim \text{Poisson}(\lambda)$ si su función de probabilidad es

$$f(x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

Esta distribución fue descubierta por Simeón Denis Poisson en 1873 como límite de la distribución binomial, al respecto véase el ejercicio 223. Puede demostrarse que $E(X) = \lambda$, y $\text{Var}(X) = \lambda$. La gráfica de la función de probabilidad Poisson se muestra en la Figura 2.14.

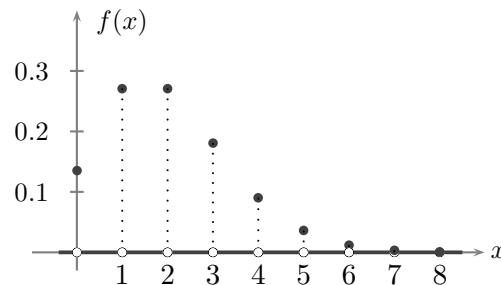


Figura 2.14: Función de probabilidad Poisson(λ) con $\lambda = 2$.

DISTRIBUCIÓN BINOMIAL NEGATIVA. Suponga que se tiene una sucesión infinita de ensayos independientes Bernoulli en donde la probabilidad de éxito en cada ensayo es $p \in (0, 1)$. Sea X el número de fracasos antes de obtener el r -ésimo éxito. Se dice entonces que X tiene una distribución binomial negativa con parámetros r y p . Se escribe $X \sim \text{bin neg}(r, p)$, y su

función de probabilidad es

$$f(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si } x = 0, 1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

Se puede demostrar que $E(X) = r(1-p)/p$, y $\text{Var}(X) = r(1-p)/p^2$. Es claro que esta distribución es una generalización de la distribución geométrica, la cual se obtiene cuando el parámetro r toma el valor 1. Para $r = 3$ y $p = 0.2$, la función de probabilidad binomial negativa tiene la forma como en la Figura 2.15.

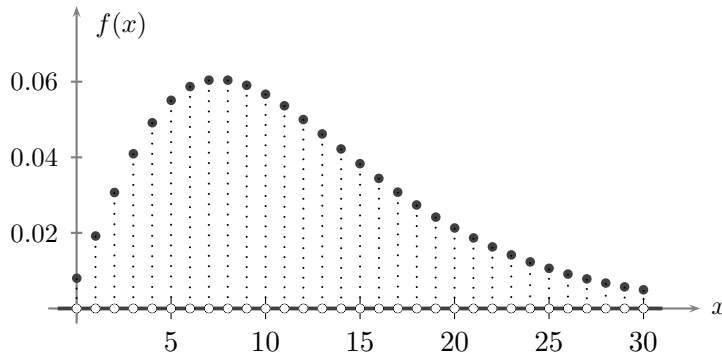


Figura 2.15: Función de probabilidad bin neg(r, p) con $r = 3$ y $p = 0.2$.

DISTRIBUCIÓN HIPERGEOMÉTRICA. Suponga que se tiene un conjunto de N objetos de los cuales K son de una primera clase, y $N - K$ son de una segunda clase. Suponga que de este conjunto se toma una muestra de tamaño n , sin reemplazo y en donde el orden de los objetos seleccionados no importa. Se define X como el número de objetos de la primera clase contenidos en la muestra seleccionada. Entonces X puede tomar los valores $0, 1, \dots, n$, suponiendo $n \leq K$. Decimos que X tiene una distribución hipergeométrica con parámetros N, K y n , se escribe $X \sim \text{hipergeo}(N, K, n)$, y su función de probabilidad es

$$f(x) = \begin{cases} \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} & \text{si } x = 0, 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}$$

La gráfica de esta función se muestra en la Figura 2.16. Es posible comprobar que

$$E(X) = n \frac{K}{N},$$

$$\text{y } \text{Var}(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}.$$

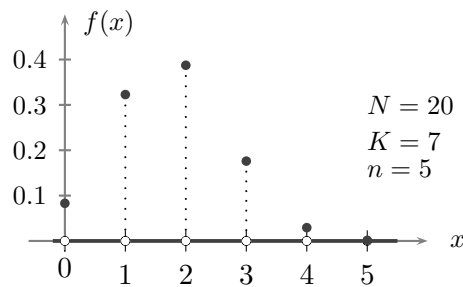


Figura 2.16: Función de probabilidad hipergeo(N, K, n).

2.7. Distribuciones continuas

Ahora se estudian algunas distribuciones de probabilidad de variables aleatorias absolutamente continuas. Algunas otras distribuciones continuas que surgen en la estadística serán estudiadas en el Capítulo 5.

DISTRIBUCIÓN UNIFORME CONTINUA. La variable aleatoria X tiene distribución uniforme en el intervalo (a, b) y se escribe $X \sim \text{unif}(a, b)$, cuando su función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b), \\ 0 & \text{otro caso.} \end{cases}$$

En este caso es inmediato verificar que $E(X) = (a+b)/2$, y $\text{Var}(X) = (b-a)^2/12$. La gráfica de esta función de densidad se muestra en la Figura 2.17

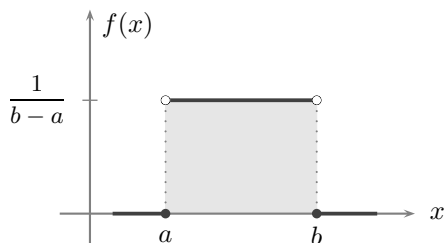


Figura 2.17: Función de densidad $\text{unif}(a, b)$.

DISTRIBUCIÓN EXPONENCIAL. La variable continua X tiene una distribución exponencial con parámetro $\lambda > 0$ y se escribe $X \sim \text{exp}(\lambda)$ cuando tiene función de densidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Para esta distribución es muy sencillo verificar que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$. Su gráfica se muestra en la Figura 2.18.

DISTRIBUCIÓN GAMA. La variable aleatoria continua X tiene distribución

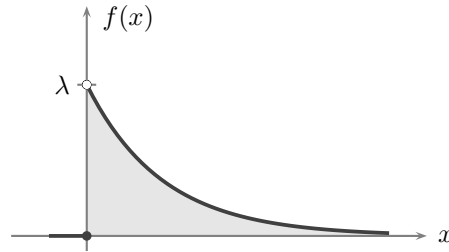


Figura 2.18: Función de densidad $\exp(\lambda)$.

gama con parámetros $n > 0$ y $\lambda > 0$ si su función de densidad es

$$f(x) = \begin{cases} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

En tal caso se escribe $X \sim \text{gama}(n, \lambda)$. La gráfica de esta función se muestra en la Figura 2.19. El término $\Gamma(n)$ es la *función gama* definida como sigue

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt,$$

para valores de n tal que la integral es convergente. Esta función satisface las siguientes propiedades:

- a) $\Gamma(n + 1) = n\Gamma(n)$.
- b) $\Gamma(n + 1) = n!$ para n entero positivo.
- c) $\Gamma(2) = \Gamma(1) = 1$.
- d) $\Gamma(1/2) = \sqrt{\pi}$.

Observe que cuando el parámetro n toma el valor 1, la distribución $\text{gama}(n, \lambda)$ se reduce a la distribución $\text{exp}(\lambda)$. Resolviendo un par de integrales se puede demostrar que $E(X) = n/\lambda$, y $\text{Var}(X) = n/\lambda^2$.

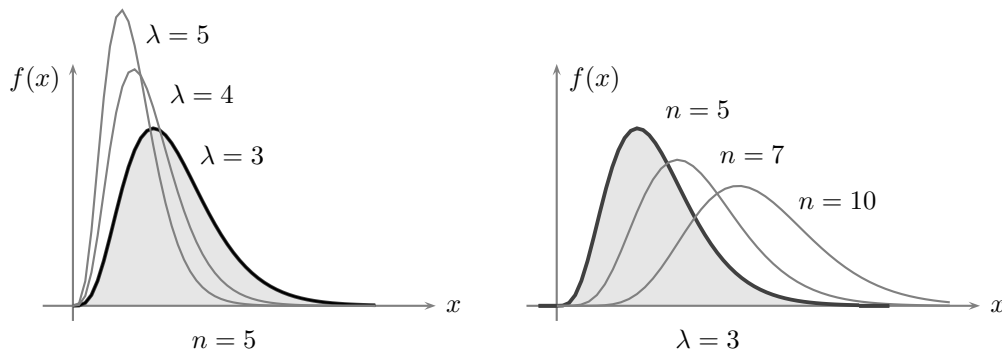


Figura 2.19: Función de densidad $\text{gama}(n, \lambda)$.

NOTA. La terminología usada para esta distribución no es estándar. En algunos otros textos aparece como $\text{gama}(\lambda, n)$, es decir, los parámetros son los mismos pero se presentan en el orden contrario. Puede entonces haber confusión cuando se escribe por ejemplo $\text{gama}(2, 3)$. En ocasiones se usa el parámetro $1/\theta$ en lugar de λ .

DISTRIBUCIÓN BETA. La variable continua X tiene distribución beta con parámetros $a > 0$ y $b > 0$, y se escribe $X \sim \text{beta}(a, b)$ cuando su función de densidad es

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

En la Figura 2.20 se ilustra la forma de esta función para varios valores de los parámetros. El término $B(a, b)$ se conoce como la *función beta*, y se

define para $a > 0$ y $b > 0$ como sigue

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Esta función satisface las siguientes propiedades.

- a) $B(a, b) = B(b, a)$.
- b) $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

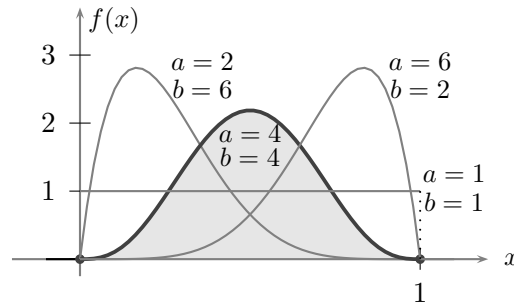


Figura 2.20: Función de densidad beta(a, b).

Por simetría se tiene que si X tiene distribución beta(a, b), entonces $1 - X$ tiene distribución beta(b, a). Para esta distribución se tiene que

$$E(X) = \frac{a}{a+b},$$

y

$$\text{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

DISTRIBUCIÓN NORMAL. Esta es posiblemente la distribución de probabilidad de mayor importancia. Se dice que la variable aleatoria continua X tiene una distribución normal o Gaussiana si su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2},$$

en donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$ son dos parámetros. En este caso se escribe $X \sim N(\mu, \sigma^2)$. No es difícil demostrar que $E(X) = \mu$, y $\text{Var}(X) = \sigma^2$. La gráfica de la función de densidad normal aparece en la Figura 2.21, en ella se muestra el significado geométrico de los parámetros. Cuando se hacen variar estos parámetros la función de densidad cambia como se ilustra en la Figura 2.22.

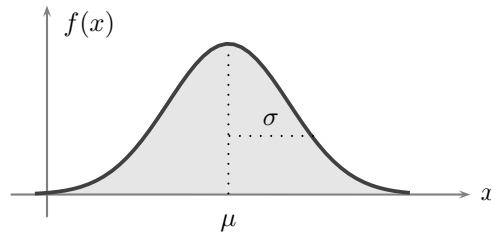


Figura 2.21: Función de densidad $N(\mu, \sigma^2)$.

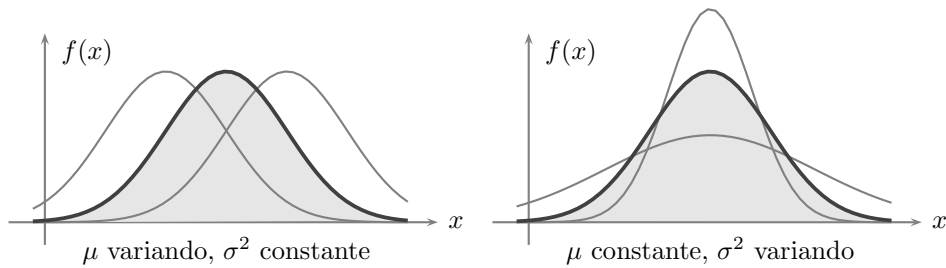


Figura 2.22: Función de densidad $N(\mu, \sigma^2)$ variando los parámetros.

En particular se dice que X tiene una distribución *normal estándar* si $\mu = 0$ y $\sigma^2 = 1$. En este caso particular, la función de densidad se reduce a la expresión más sencilla

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Es posible transformar una variable aleatoria normal no estándar en una

estándar mediante la siguiente operación llamada *estandarización*. La demostración de este resultado es elemental y se deja como ejercicio.

$$\text{PROPOSICIÓN. } X \sim N(\mu, \sigma^2) \iff Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Comúnmente se usa la letra Z para denotar una variable aleatoria con distribución normal estándar. En particular la función $\Phi(x)$ denota la función de distribución de una variable aleatoria normal estándar, es decir,

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

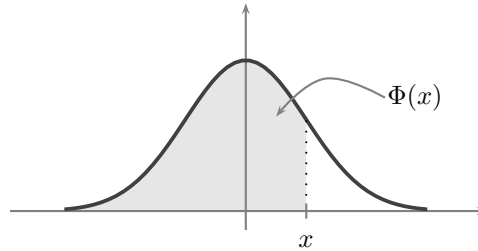


Figura 2.23: Área cubierta por la función de distribución $\Phi(x) = P(Z \leq x)$.

Los valores de esta función no pueden encontrarse de manera explícita, así es que se usan métodos numéricos para aproximar la integral para distintos valores de x . En una tabla al final del texto pueden encontrarse estos valores aproximados.

DISTRIBUCIÓN LOG NORMAL. Si X tiene distribución $N(\mu, \sigma^2)$, entonces la variable $Y = e^X$ tiene una distribución log normal (μ, σ^2) , y su función de densidad es

$$f(y) = \begin{cases} \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

La gráfica de esta función de densidad se muestra en la Figura 2.24. Se puede demostrar que

$$\begin{aligned} E(Y) &= \exp(\mu + \sigma^2/2), \\ \text{y } \text{Var}(Y) &= \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2). \end{aligned}$$

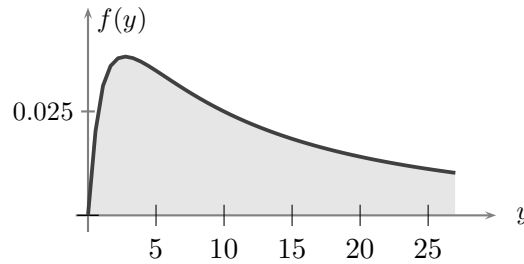


Figura 2.24: Función de densidad log normal(μ, σ^2) con $\mu = 3$ y $\sigma^2 = 2$.

Algunas otras distribuciones continuas de interés se encuentran en el capítulo sobre distribuciones muestrales.

2.8. Ejercicios

VARIABLES ALEATORIAS

91. Demuestre que la función identidad $X(\omega) = \omega$ no es variable aleatoria cuando $\Omega = \{1, 2, 3\}$ y $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$.
92. Sea $\Omega = \{-1, 0, 1\}$ y $\mathcal{F} = \{\emptyset, \{0\}, \{-1, 1\}, \Omega\}$. Considere la función identidad $X(\omega) = \omega$. Demuestre que X^2 es variable aleatoria pero X no lo es.
93. Considere el espacio medible (Ω, \mathcal{F}) , con $\mathcal{F} = \{\emptyset, \Omega\}$. Demuestre que la función $X : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, X es constante.
94. Sea (Ω, \mathcal{F}) un espacio medible tal que $\mathcal{F} = \{\emptyset, \Omega, A, A^c\}$ con $A \subseteq \Omega$. Demuestre que toda función medible $X : \Omega \rightarrow \mathbb{R}$ es constante en A y en A^c . Por lo tanto toda función medible respecto de esta σ -álgebra toma a lo sumo dos valores distintos. El siguiente ejercicio generaliza este resultado.
95. Sea A_1, \dots, A_n una partición finita de Ω , y considere el espacio medible (Ω, \mathcal{F}) , con $\mathcal{F} = \sigma\{A_1, \dots, A_n\}$. Demuestre que $X : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, X es constante en cada elemento de la partición. En consecuencia, X toma a lo sumo n valores distintos.
96. Demuestre que X es variable aleatoria si, y sólo si, $(X < x) \in \mathcal{F}$ para cada número real x .
97. Demuestre que X es variable aleatoria si, y sólo si, $(X \geq x) \in \mathcal{F}$ para cada número real x .
98. Demuestre que X es variable aleatoria si, y sólo si, $(X > x) \in \mathcal{F}$ para cada número real x .
99. Demuestre que X es variable aleatoria si, y sólo si, $(a < X < b) \in \mathcal{F}$ para cada intervalo (a, b) de \mathbb{R} .

100. Sea c una constante y X una variable aleatoria. Demuestre directamente que las siguientes funciones también son variables aleatorias: cX , $X + c$, $\max\{X, c\}$, $\min\{X, c\}$.
101. Demuestre directamente que la diferencia de dos variables aleatorias es variable aleatoria.
102. Sea X una variable aleatoria cualquiera. Demuestre que la parte entera de X , denotada por $\lfloor X \rfloor$, es una variable aleatoria discreta, es decir, toma un número numerable de valores.
103. Demuestre que el conjunto de variables aleatorias definidas sobre un espacio de probabilidad es un espacio vectorial con las operaciones usuales de suma y producto por escalares.
104. Sean X y Y variables aleatorias. Demuestre directamente que tanto $\max\{X, Y\}$ como $\min\{X, Y\}$ son variables aleatorias.
105. Demuestre directamente que si X es variable aleatoria, entonces también lo son X^n y $2X^3 - 5X$.
106. Demuestre que X es variable aleatoria si, y sólo si, tanto $X^+ = \max\{0, X\}$ como $X^- = -\min\{0, X\}$, lo son.
107. Sea $A \subseteq \Omega$. Demuestre que la función indicadora $1_A : \Omega \rightarrow \mathbb{R}$ es variable aleatoria si, y sólo si, el conjunto A es medible. Véase el apéndice al final del texto para la definición y algunas propiedades de la función indicadora.
108. Sean $A, B \subseteq \Omega$. Diga falso o verdadero. Demuestre en cada caso.
- a) A, B medibles $\Rightarrow 1_A + 1_B$ es v.a.
- b) $1_A + 1_B$ es v.a. $\Rightarrow A, B$ son medibles.
109. Sean A, B subconjuntos disjuntos de Ω y sean a, b dos números reales distintos. Demuestre que

$$a1_A + b1_B \text{ es v.a.} \Leftrightarrow A, B \text{ son medibles.}$$

Una de estas implicaciones resulta falsa cuando se omite la condición de que los números a y b son distintos. ¿Cuál de ellas es?

110. Sean A_1, \dots, A_n subconjuntos disjuntos de Ω , y sean a_1, \dots, a_n constantes distintas. Demuestre que

$$\sum_{i=1}^n a_i 1_{A_i} \text{ es v.a.} \Leftrightarrow A_1, \dots, A_n \text{ son medibles.}$$

111. Sean A y B dos eventos, y sean 1_A y 1_B las correspondientes funciones indicadoras. Directamente de la definición demuestre que las funciones $1_A + 1_B$, $1_A - 1_B$ y $1_A 1_B$ son variables aleatorias.
112. Sean X y Y dos variables aleatorias. Demuestre que los conjuntos $(X \leq Y)$, $(X = Y)$, $(X - Y < 1)$, $(X - Y > 0)$, $(X \geq Y)$ y $(X \neq Y)$ son eventos.
113. Sean X , Y y Z tres variables aleatorias. Demuestre que los conjuntos $(X \leq Y \leq Z)$, $(X = Y = Z)$ y $(X > Y > Z)$ son eventos.
114. Sea X una variable aleatoria y $g : \mathbb{R} \rightarrow \mathbb{R}$ una función Borel medible. Demuestre que $g(X) = g \circ X : \Omega \rightarrow \mathbb{R}$ es también una variable aleatoria. Sugerencia: Demuestre que la colección $\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}) : g^{-1}B \in \mathcal{B}(\mathbb{R})\}$ coincide con $\mathcal{B}(\mathbb{R})$ usando los siguientes dos resultados: (1) Dada una función continua de \mathbb{R} en \mathbb{R} , la imagen inversa de un conjunto abierto es nuevamente un conjunto abierto. (2) Todo conjunto abierto de \mathbb{R} distinto del vacío puede expresarse como una unión numerable de intervalos abiertos.
115. Sea X una variable aleatoria. Demuestre que las funciones e^X , $\text{sen } X$, y $\text{cos } X$ son variables aleatorias.
116. Sea $X : \Omega \rightarrow \mathbb{R}$ una función. Proporcione un ejemplo en el que X^2 es variable aleatoria pero $|X|$ no lo es.
117. Sean X_1, \dots, X_n variables aleatorias. Demuestre que

$$a) \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{es v.a.}$$

$$b) S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{es v.a.}$$

118. Sea X una variable aleatoria, y sean $a < b$ dos constantes. Demuestre que las siguientes funciones son variables aleatorias.

$$a) Y = \begin{cases} X & \text{si } X < a, \\ a & \text{si } X \geq a. \end{cases}$$

$$b) Y = \begin{cases} a & \text{si } X < a, \\ X & \text{si } a \leq X \leq b, \\ b & \text{si } X > b. \end{cases}$$

$$c) Y = \begin{cases} X & \text{si } |X| \leq a, \\ 0 & \text{si } |X| > a, \end{cases} \quad \text{suponiendo } a > 0.$$

119. Se define la función signo como sigue

$$\text{signo}(x) = \begin{cases} +1 & \text{si } x > 0, \\ -1 & \text{si } x < 0, \\ 0 & \text{si } x = 0. \end{cases}$$

Demuestre que si X es variable aleatoria, entonces $\text{signo}(X)$ también lo es. ¿Es cierto el recíproco?

120. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sea $X : \Omega \rightarrow \mathbb{R}$ una función. Demuestre que la colección $\{X^{-1}B : B \in \mathcal{B}(\mathbb{R})\}$ es una sub σ -álgebra de \mathcal{F} si, y sólo si, X es variable aleatoria. A esta colección se le denota por $\sigma(X)$, y es la mínima σ -álgebra respecto de la cual X es variable aleatoria.

121. Sea X una variable aleatoria con valores en el conjunto $\{0, 1, \dots\}$. Sea $(X)_{10}$ el valor de X módulo 10. Demuestre que $(X)_{10}$ es también variable aleatoria.

122. MEDIDA DE PROBABILIDAD INDUCIDA. Sean $(\Omega_1, \mathcal{F}_1)$ y $(\Omega_2, \mathcal{F}_2)$ dos espacios medibles, y sea $X : \Omega_1 \rightarrow \Omega_2$ una función medible, es decir, para cualquier A en \mathcal{F}_2 se cumple que $X^{-1}A \in \mathcal{F}_1$. Suponga que $P : \mathcal{F}_1 \rightarrow [0, 1]$ es una medida de probabilidad. Demuestre que $P \circ X^{-1} : \mathcal{F}_2 \rightarrow [0, 1]$ es también una medida de probabilidad. A esta función se le llama medida de probabilidad inducida por X .
123. Sea c una constante distinta de cero, y sea X una variable aleatoria. Demuestre o proporcione un contraejemplo.
- $\sigma(cX) = \sigma(X)$.
 - $\sigma(X + c) = \sigma(X)$.
 - $\sigma(X) = \sigma(X^2)$.

Función de distribución

124. Grafique y demuestre que las siguientes funciones son de distribución.

$$a) F(x) = \begin{cases} 1 - e^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

$$b) F(x) = \begin{cases} 1 - (1+x)e^{-x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

$$c) F(x) = \begin{cases} 0 & \text{si } x < -1, \\ (x+1)/2 & \text{si } x \in [-1, 1], \\ 1 & \text{si } x > 1. \end{cases}$$

125. Investigue si las siguientes funciones son de distribución.

$$a) F(x) = \begin{cases} 1 - e^{-x^2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

$$b) F(x) = \begin{cases} e^{-1/x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

$$c) F(x) = e^x / (1 + e^x), \text{ para } x \in \mathbb{R}.$$

$$d) F(x) = e^x / (e^x + e^{-x}), \text{ para } x \in \mathbb{R}.$$

126. Sean $F(x)$ y $G(x)$ dos funciones de distribución. Determine si las siguientes funciones son de distribución.

a) $aF(x) + (1 - a)G(x)$, con $0 \leq a \leq 1$.

b) $F(x) + G(x)$.

c) $F(x)G(x)$.

d) $\frac{2G(x)}{1 + F(x)}$.

127. Sea X con función de distribución la especificada abajo. Grafique $F(x)$ y demuestre que es efectivamente una función de distribución. Calcule además $P(X \leq 4)$, $P(X > 1)$, $P(4 < X < 6)$ y $P(X = 2)$.

$$F(x) = \begin{cases} 0 & \text{si } x < 2, \\ 1 - 4/x^2 & \text{si } x \geq 2. \end{cases}$$

128. Sea X con función de distribución la especificada abajo. Grafique $F(x)$ y demuestre que es efectivamente una función de distribución. Calcule además $P(X \leq 1)$, $P(X = 1)$, $P(0 < X < 3)$, $P(X = 4)$ y $P(X \geq 3)$.

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 0.2 & \text{si } 0 \leq x < 1, \\ 0.5 & \text{si } 1 \leq x < 3, \\ 0.9 & \text{si } 3 \leq x < 4, \\ 1 & \text{si } x \geq 4. \end{cases}$$

129. En la escuela rusa de probabilidad se define la función de distribución de una variable aleatoria X como $G(x) = P(X < x)$. Observe el signo “<” en lugar de “ \leq ”. Demuestre que esta función cumple todas las propiedades de una función de distribución, excepto que ahora la continuidad es por la izquierda.

130. Sea $F(x)$ una función de distribución continua. Demuestre que para cualquier entero $n \geq 1$, las siguientes funciones también son de distribución.

- a) $[F(x)]^n$.
 b) $1 - [1 - F(x)]^n$.
131. Sea X con función de distribución $F(x)$. Diga falso o verdadero, demuestre en cada caso. Para todo $x \in \mathbb{R}$,
- a) $F(x) = P(X < x) + P(X = x)$.
 b) $1 - F(x) = P(X \geq x)$.
 c) $1 - P(X < x) - P(X > x) = P(X = x)$.
 d) $F(x) - \frac{1}{2}P(X = x) = \frac{1}{2}(F(x) + F(x-))$.
132. Encuentre la función de distribución de la variable Y en términos de la función de distribución de X cuando
- a) $Y = aX + b$, con a, b constantes. f) $Y = X^- = -\min\{0, X\}$.
 b) $Y = e^X$. g) $Y = |X|$.
 c) $Y = e^{-X}$. h) $Y = -X$.
 d) $Y = X^2$. i) $Y = \text{sen } X$.
 e) $Y = X^+ = \max\{0, X\}$. j) $Y = \cos X$.
133. Sea X con función de distribución $F_X(x)$, y sean $a < b$ dos constantes. Calcule la función de distribución de Y en términos de la función de distribución de X , y muestre gráficamente el comportamiento de $F_Y(y)$ en los puntos a y b .
- a) $Y = \begin{cases} X & \text{si } X < a, \\ a & \text{si } X \geq a. \end{cases}$
 b) $Y = \begin{cases} a & \text{si } X < a, \\ X & \text{si } a \leq X \leq b, \\ b & \text{si } X > b. \end{cases}$
 c) $Y = \begin{cases} X & \text{si } |X| \leq a, \\ 0 & \text{si } |X| > a, \end{cases}$ con $a > 0$.
134. Sean $F(x)$ y $G(x)$ dos funciones de distribución continuas y estrictamente crecientes. Demuestre que

- a) si $F(x) \geq G(x)$, entonces $F^{-1}(y) \leq G^{-1}(y)$.
- b) si X tiene función de distribución $F(x)$, entonces $Y = G^{-1}(F(X))$ tiene función de distribución $G(x)$.
- c) si $F(x) \geq G(x)$, entonces existen variables aleatorias X y Y cuyas funciones de distribución son $F(x)$ y $G(x)$ respectivamente, y son tales que $X \leq Y$. Sugerencia: Use el inciso anterior.
135. Sea X con función de distribución $F(x)$. Demuestre que $F(x)$ es continua en $x = x_0$ si, y sólo si, $P(X = x_0) = 0$.

Tipos de variables aleatorias

136. Encuentre la constante c que hace a $f(x)$ una función de probabilidad.
- a) $f(x) = \frac{c}{x(x+1)}$, para $x = 1, 2, \dots$
- b) $f(x) = ce^{-x}$, para $x = 1, 2, \dots$
- c) $f(x) = c/x!$, para $x = 1, 2, \dots$
137. Encuentre la constante c que hace a $f(x)$ una función de densidad.
- a) $f(x) = cx^2$, para $0 < x < 1$.
- b) $f(x) = cxe^{-2x^2}$, para $x > 0$.
- c) $f(x) = cx^{-2}$, para $x > 1$.
- d) $f(x) = \frac{ce^x}{(1+e^x)^2}$, para $x \in \mathbb{R}$.
- e) $f(x) = cx(1-x)$, para $0 < x < 1$.
- f) $f(x) = \frac{c}{\sqrt{1-x^2}}$, para $0 < x < 1$.
- g) $f(x) = \frac{c}{1+x^2}$, para $x \in \mathbb{R}$.
138. Demuestre que las siguientes funciones son de densidad. Encuentre la correspondiente función de distribución y demuestre que ésta es efectivamente una función de distribución. Grafique ambas funciones.

- a) $f(x) = 2x$, para $x \in (0, 1)$.
- b) $f(x) = 3x^2/2$, para $x \in (-1, 1)$.
- c) $f(x) = 1 - x/2$, para $x \in (0, 2)$.
- d) $f(x) = 2x/m^2$, para $x \in (0, m)$, con $m > 0$.
- e) $f(x) = 1/(1 - x)^2$, para $x \in (0, 1/2)$.
- f) $f(x) = e^{|x|}/2$, para $x \in \mathbb{R}$.

139. Demuestre que las siguientes funciones son de distribución. Encuentre la correspondiente función de densidad y compruebe que ésta es efectivamente una función de densidad. Grafique ambas funciones.

- a) $F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$
- b) $F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$
- c) $F(x) = e^x/(1 + e^x)$.
- d) $F(x) = \frac{1}{2} \int_{-\infty}^x e^{-|u|} du$.

140. Sea $f(x)$ una función de densidad y sea c una constante cualquiera. Demuestre que $f(x + c)$ es también una función de densidad.

141. Diga falso o verdadero. Demuestre en cada caso.

- a) Toda función de densidad es acotada.
- b) Toda función de distribución es acotada.

142. Sea X absolutamente continua, y sea $Y = aX + b$ con a y b constantes. Demuestre que si $a \neq 0$, entonces

$$f_Y(y) = \frac{1}{|a|} f_X((y - b)/a).$$

Igualdad de variables aleatorias

143. Demuestre que la igualdad casi segura de variables aleatorias es una relación de equivalencia. ¿Cumple tal propiedad la igualdad en distribución?
144. Sea $X \geq 0$ tal que $E(X) = 0$. Demuestre que $X = 0$ c.s. Sugerencia: Para cada natural n defina el evento $A_n = (X \geq 1/n)$. Compruebe que $E(X) \geq E(X \cdot 1_{A_n}) \geq P(A_n)/n$. Esto lleva a la conclusión de que $P(A_n) = 0$ y por lo tanto $P(\cup_{n=1}^{\infty} A_n) = 0$. Ahora compruebe que los eventos $(X > 0)$ y $\cup_{n=1}^{\infty} A_n$ coinciden. Alternativamente puede usarse la desigualdad de Markov (ver página 347).

Integral de Riemann-Stieltjes

145. Sea X una variable aleatoria con función de distribución F , y sea a cualquier número real. Demuestre que

$$\int_{-\infty}^{\infty} 1_{\{a\}}(x) dF(x) = P(X = a).$$

146. Sea X una variable aleatoria con función de distribución F , y sea $(a, b) \subseteq \mathbb{R}$. Demuestre que

$$\int_{-\infty}^{\infty} 1_{(a,b)}(x) dF(x) = P(a < X < b).$$

147. Sea F una función de distribución absolutamente continua. Demuestre que para cualesquiera números naturales n y m ,

$$\int_{-\infty}^{\infty} F^n(x) dF^m(x) = \frac{m}{n+m}.$$

Esperanza

148. Calcule la esperanza de X cuya función de probabilidad o de densidad es

a) $f(x) = 1/5$, para $x = -2, -1, 0, 1, 2$.

b) $f(x) = e^{-1}/x!$, para $x = 0, 1, 2, \dots$

c) $f(x) = |x|$, para $-1 < x < 1$.

d) $f(x) = e^{-|x|}/2$, para $x \in \mathbb{R}$.

149. Calcule la esperanza de una variable aleatoria cuya función de distribución es

$$F(x) = \begin{cases} 1 - e^{-x}/2 & \text{si } x > 1, \\ 0 & \text{si } x \leq 1. \end{cases}$$

150. Sean X y Y con esperanza finita, y sea c una constante. Demuestre que

a) $E(c) = c$.

b) $E(cX) = cE(X)$.

c) $E(X + c) = E(X) + c$.

d) Si $X \geq 0$, entonces $E(X) \geq 0$.

e) Si $X \leq Y$, entonces $E(X) \leq E(Y)$.

f) $|E(X)| \leq E|X|$.

151. Demuestre que no existe la esperanza de X cuando su función de probabilidad o de densidad es

a) $f(x) = \frac{3}{\pi^2 x^2}$, para $x \in \mathbb{Z} \setminus \{0\}$.

b) $f(x) = \frac{1}{\pi(1+x^2)}$, para $x \in \mathbb{R}$.

152. LA PARADOJA DE SAN PETERSBURGO. Un juego consiste en lanzar una moneda equilibrada repetidas veces hasta que una de las caras, seleccionada previamente, aparezca por primera vez. Si un jugador lanza la moneda y requiere de n lanzamientos para que se cumpla la condición, entonces recibe 2^n unidades monetarias. ¿Cuál debe ser el pago inicial justo para ingresar a este juego?
153. Sea $\{A_1, A_2, \dots\}$ una colección de eventos que forman una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva. Sea X una variable aleatoria discreta con esperanza finita. Para cualquier evento A con probabilidad positiva defina

$$E(X | A) = \sum_x xP(X = x | A).$$

Demuestre que
$$E(X) = \sum_{i=1}^{\infty} E(X | A_i)P(A_i).$$

154. Sean X y Y con esperanza finita. Demuestre que
- $E(\min\{X, Y\}) \leq \min\{E(X), E(Y)\} \leq E(X)$.
 - $E(\max\{X, Y\}) \geq \max\{E(X), E(Y)\} \geq E(X)$.
155. Sea $X > 0$, discreta y con esperanza finita. Demuestre directamente que $E(X)E(1/X) \geq 1$. Este resultado puede ser demostrado usando la desigualdad de Jensen (ver página 127), pero en este ejercicio se pide obtener el resultado sin usar dicha desigualdad.
156. Sea X discreta con valores no negativos $x_1 \leq x_2 \leq \dots \leq x_k$. Demuestre que

$$a) \lim_{n \rightarrow \infty} \frac{E(X^{n+1})}{E(X^n)} = x_k,$$

$$b) \lim_{n \rightarrow \infty} \sqrt[n]{E(X^n)} = x_1.$$

157. Sea X discreta con valores $0, 1, \dots$ y con esperanza finita. Demuestre que

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} P(X > n).$$

Use esta fórmula para demostrar que

- a) si X tiene distribución $\text{geo}(p)$, entonces $E(X) = (1-p)/p$.
 b) si X tiene distribución $\text{Poisson}(\lambda)$, entonces $E(X) = \lambda$.
158. Sea $X \geq 0$ con esperanza finita, y suponga que para algún $p \in (0, 1)$, se cumple la desigualdad $P(X \geq k) \leq p^k$, para cada $k = 0, 1, \dots$. Demuestre que $E(X) \leq 1/(1-p)$.

159. Sea $X \geq 0$ con esperanza finita no necesariamente discreta. Para cada número natural n defina el evento $A_n = (n-1 \leq X < n)$. Demuestre que

$$\sum_{n=1}^{\infty} (n-1)1_{A_n} \leq X < \sum_{n=1}^{\infty} n1_{A_n}.$$

Ahora demuestre las desigualdades

$$\sum_{n=1}^{\infty} P(X \geq n) \leq E(X) < 1 + \sum_{n=1}^{\infty} P(X \geq n).$$

160. Sea X con función de distribución $F(x)$. Demuestre que si X tiene esperanza finita, entonces

- a) $\lim_{x \rightarrow \infty} x(1 - F(x)) = 0$.
 b) $\lim_{x \rightarrow -\infty} xF(x) = 0$.

El recíproco sin embargo es falso, véase [4].

161. Sea X con función de distribución $F(x)$, y con esperanza finita. Demuestre que

$$E(X) = \int_0^{\infty} [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

Gráficamente estas integrales pueden interpretarse como se indica en la Figura 2.25.

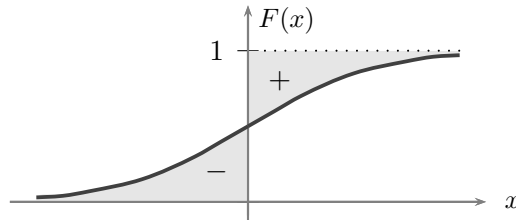


Figura 2.25: La esperanza como la diferencia de dos áreas.

Use esta fórmula para demostrar que

- a) si X tiene distribución $\exp(\lambda)$, entonces $E(X) = 1/\lambda$.
- b) si X tiene distribución $\text{gama}(n, \lambda)$, entonces $E(X) = n/\lambda$.

162. Sea X una variable aleatoria no negativa con función de distribución continua $F(x)$ y con esperanza finita μ . Demuestre que la siguiente función es de distribución.

$$G(y) = \begin{cases} 1 - \frac{1}{\mu} \int_y^\infty (1 - F(x)) dx & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

Demuestre que la esperanza de esta distribución es $2E(X^2)/\mu$, suponiendo que el segundo momento de X es finito.

163. Sea X con función de distribución continua $F(x)$, y con esperanza finita μ . Demuestre que

$$\int_{-\infty}^{\mu} F(x) dx = \int_{\mu}^{\infty} [1 - F(x)] dx.$$

164. Demuestre que la condición $E(X) = 0$ no implica que X es simétrica alrededor de cero. Sugerencia: Considere X tal que $P(X = -1) = 1/2$,

$P(X = 0) = 1/8$, $P(X = 1) = 1/4$ y $P(X = 2) = 1/8$. ¿Puede usted construir un ejemplo de una distribución continua con esperanza cero, que no sea simétrica?

165. Calcule la esperanza de una variable aleatoria con función de distribución continua dada por la gráfica de la Figura 2.26. Calcule y grafique además la correspondiente función de densidad.

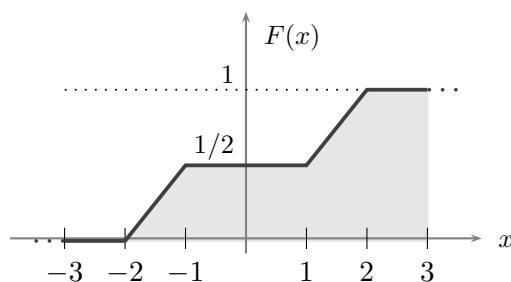


Figura 2.26: Una función de distribución continua.

166. Calcule la esperanza de una variable aleatoria con función de distribución dada por la gráfica de la Figura 2.27.

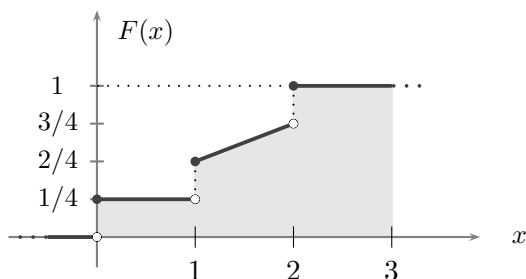


Figura 2.27: Una función de distribución mixta.

167. Demuestre que si $X = 0$ c.s., entonces $E(X) = 0$.
168. Sean X y Y con esperanza finita tales que $X = Y$ c.s. Demuestre que $E(X) = E(Y)$.

Varianza

169. Calcule la varianza de X cuya función de probabilidad o de densidad es

a) $f(x) = 1/5$, para $x = -2, -1, 0, 1, 2$.

b) $f(x) = e^{-1}/x!$, para $x = 0, 1, 2, \dots$

c) $f(x) = |x|$, para $-1 < x < 1$.

d) $f(x) = e^{-|x|}/2$, para $x \in \mathbb{R}$.

170. Sean X y Y con varianza finita y sea c una constante. Demuestre las siguientes propiedades de la varianza.

a) $\text{Var}(X) \geq 0$.

b) $\text{Var}(cX) = c^2 \text{Var}(X)$.

c) $\text{Var}(X + c) = \text{Var}(X)$.

d) $\text{Var}(X) = E(X^2) - E^2(X)$.

171. Use la desigualdad de Chebyshev para demostrar que $\text{Var}(X) = 0$ si, y sólo si, X es constante.

172. Sea X con valores en $[a, b]$. Demuestre que

a) $a \leq E(X) \leq b$.

b) $0 \leq \text{Var}(X) \leq (b - a)^2/4$.

173. MINIMIZACIÓN DEL ERROR CUADRÁTICO MEDIO. Sea X con segundo momento finito. A la función $g(u) = E[(X - u)^2]$ se le conoce como error cuadrático medio. Demuestre que $g(u)$ se minimiza cuando $u = E(X)$. En consecuencia, para cualquier valor real de u ,

$$\text{Var}(X) \leq E[(X - u)^2].$$

174. Sea X con varianza finita y sea c una constante. Demuestre que

$$E(X - c)^2 = \text{Var}(X) + (E(X) - c)^2.$$

175. Sea X con media μ y varianza σ^2 . Demuestre que $E|X - \mu| \leq \sigma$.
Sugerencia: $\text{Var}(|X - \mu|) \geq 0$.
176. Diga falso o verdadero. Demuestre en cada caso.
- Si $X \leq Y$, entonces $\text{Var}(X) \leq \text{Var}(Y)$.
 - $\text{Var}(X) \leq E(X^2)$.
 - $E^2(X) \leq E(X^2)$.
177. Sea X una variable aleatoria con varianza finita, y sea a una constante. Diga si las siguientes afirmaciones son falsas o verdaderas, demuestre en cada caso.
- $E(\text{mín}\{X, a\}) \leq E(X) \leq E(\text{máx}\{X, a\})$.
 - $\text{Var}(\text{mín}\{X, a\}) \leq \text{Var}(X) \leq \text{Var}(\text{máx}\{X, a\})$.
178. Sean X y Y con varianza finita. Diga si las siguientes desigualdades son falsas o verdaderas, demuestre en cada caso.
- $\text{Var}(\text{mín}\{X, Y\}) \leq \text{Var}(X) \leq \text{Var}(\text{máx}\{X, Y\})$.
 - $\text{Var}(X + Y) \leq 2(\text{Var}(X) + \text{Var}(Y))$.
 - $\sqrt{\text{Var}(X + Y)} \leq \sqrt{\text{Var}(X)} + \sqrt{\text{Var}(Y)}$.
179. Sea X con varianza finita, y sea c una constante cualquiera. Diga si las siguientes afirmaciones son falsas o verdaderas, demuestre en cada caso.
- $\text{Var}(X + c) = \text{Var}(X - c)$.
 - $\text{Var}(|X|) \leq \text{Var}(X)$.
 - $\text{Var}(|X - c|) \leq \text{Var}(X)$.
180. Calcule la varianza de una variable aleatoria cuya función de distribución está dada por la gráfica de la Figura 2.28.
181. Sean X y Y independientes y con segundo momento finito. Demuestre que
- $$\text{Var}(XY) = \text{Var}(X)\text{Var}(Y) + E^2(X)\text{Var}(Y) + E^2(Y)\text{Var}(X).$$

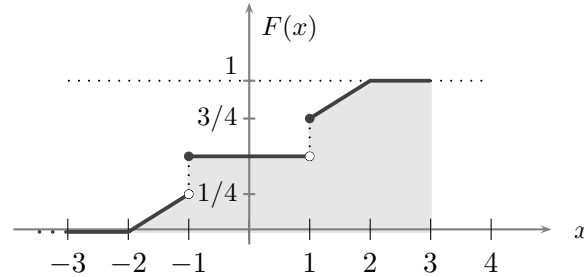


Figura 2.28: Una función de distribución mixta.

182. Sean X y Y con segundo momento finito. Demuestre que

$$|\sqrt{\text{Var}(X)} - \sqrt{\text{Var}(Y)}| \leq \sqrt{\text{Var}(X \pm Y)} \leq \sqrt{\text{Var}(X)} + \sqrt{\text{Var}(Y)}.$$

Momentos

183. Calcule el n -ésimo momento de una variable aleatoria cuya función de probabilidad o de densidad es

- a) $f(x) = 1/5$, para $x = -2, -1, 0, 1, 2$.
- b) $f(x) = e^{-1}/x!$, para $x = 0, 1, 2, \dots$
- c) $f(x) = |x|$, para $-1 < x < 1$.
- d) $f(x) = e^{-|x|}/2$, para $x \in \mathbb{R}$.

184. Sea X con n -ésimo momento finito. Demuestre que para cualquier número natural $m \leq n$, se cumple $E|X|^m \leq E|X|^n$. Este resultado establece que si el n -ésimo momento de una variable aleatoria es finito, entonces todos los momentos anteriores a n también son finitos. Sugerencia: $|X|^m = |X|^m \cdot 1_{(|X| \leq 1)} + |X|^m \cdot 1_{(|X| > 1)}$.

185. Sea X con distribución simétrica alrededor de $x = 0$, y con cuarto momento finito. Demuestre que para cualquier número real a ,

$$E(X^4) \leq E(X - a)^4.$$

186. Sea 1_A la función indicadora de un evento A . Demuestre que

$$a) E(1_A) = E(1_A^n) = P(A).$$

$$b) \text{Var}(1_A) = P(A)(1 - P(A)) \leq 1/4.$$

187. Sea X con n -ésimo momento finito. Demuestre que

$$E|X|^n = n \int_0^\infty x^{n-1} (1 - F(x)) dx + n \int_{-\infty}^0 |x|^{n-1} F(x) dx.$$

188. Sea X discreta con valores en el conjunto $\{0, 1, \dots\}$, y con segundo momento finito. Demuestre que

$$E(X^2) = \sum_{n=1}^{\infty} (2n - 1)P(X \geq n).$$

189. Sea $X \geq 0$ continua y con segundo momento finito. Demuestre que

$$E(X^2) = 2 \int_0^\infty xP(X > x) dx.$$

190. ESPACIO L^1 . Demuestre que el espacio $L^1(\Omega, \mathcal{F}, P)$ consistente de todas las variables aleatorias X tales que $E|X| < \infty$, es un espacio vectorial. Para resolver este ejercicio suponga válida la propiedad de linealidad de la esperanza. Tal propiedad será demostrada más adelante.

191. DESIGUALDAD DE CAUCHY-SCHWARZ. Sean X y Y con segundo momento finito. Demuestre que

$$E^2(XY) \leq E(X^2)E(Y^2).$$

Sugerencia: Para cualquier valor real de t , la esperanza de $(tX + Y)^2$ es no negativa. Desarrolle el cuadrado y encuentre una ecuación cuadrática en t . ¿Qué puede decir de su discriminante?

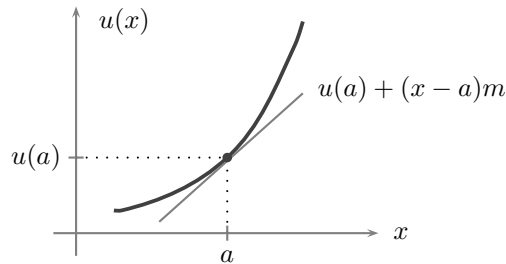


Figura 2.29: Convexidad.

192. ESPACIO L^2 . Use la desigualdad de Cauchy-Schwarz para demostrar que el espacio $L^2(\Omega, \mathcal{F}, P)$ consistente de todas las variables aleatorias X tales que $E|X|^2 < \infty$, es un espacio vectorial.
193. DESIGUALDAD DE JENSEN. Sea u una función convexa, y sea X una variable aleatoria con esperanza finita. Demuestre que

$$u(E(X)) \leq E(u(X)).$$

Sugerencia: La función u es convexa si para cada a existe un número m tal que $u(x) \geq u(a) + (x - a)m$, para todo x . Esto se muestra en la Figura 2.29. Alternativamente, una función u es convexa si $u(tx + (1 - t)y) \leq tu(x) + (1 - t)u(y)$, para cualesquiera par de números x y y dentro del dominio de definición de u , y para cualquier t en el intervalo $[0, 1]$. Debe suponerse además que el número $tx + (1 - t)y$ pertenece también al dominio de definición de la función. Vea el siguiente ejercicio para algunos ejemplos particulares de funciones convexas.

194. Sea X con esperanza finita. Use la desigualdad de Jensen para demostrar que
- $e^{E(X)} \leq E(e^X)$.
 - $E^2(X) \leq E(X^2)$.
 - $\max\{a, E(X)\} \leq E\{\max\{a, X\}\}$, a constante.

$$d) \frac{1}{E(X)} \leq E(1/X), \quad \text{suponiendo } X > 0.$$

195. Demuestre que si X es una variable aleatoria acotada casi seguramente, es decir, existe $k > 0$ tal que $P(|X| \leq k) = 1$, entonces todos los momentos de X existen.

196. Sea X una variable aleatoria con función de densidad dada por

$$f(x) = \begin{cases} n/x^{n+1} & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

Demuestre que esta función es de densidad para cualquier valor natural del parámetro n . Demuestre además que tal variable aleatoria tiene momentos finitos de orden $1, 2, \dots, n-1$, pero el n -ésimo momento y superiores no existen.

197. DESIGUALDAD c_r . Demuestre que para cada $r > 0$,

$$E|X + Y|^r \leq c_r (E|X|^r + E|Y|^r),$$

en donde c_r es una constante dada por

$$c_r = \begin{cases} 1 & \text{si } 0 < r \leq 1, \\ 2^{r-1} & \text{si } r > 1. \end{cases}$$

En particular, este resultado establece que si X y Y tienen r -ésimo momento absoluto finito, entonces $X+Y$ también. Sugerencia: A partir de la identidad $(1+t)^r = c_r (1+t^r)$, válida para cada $t \geq 0$, demuestre que para cualesquiera números reales x y y , $|x+y|^r \leq c_r (|x|^r + |y|^r)$.

198. DESIGUALDAD DE HÖLDER. Sean r y s dos números reales tales que $r > 1$ y $1/r + 1/s = 1$. Demuestre que

$$E|XY| \leq (E|X|^r)^{1/r} (E|Y|^s)^{1/s}.$$

Sugerencia: Use la desigualdad $|xy| \leq |x|^r/r + |y|^s/s$, válida para cualesquiera números reales x y y , y para r y s con las condiciones mencionadas. El caso $r = s = 2$ corresponde a la desigualdad de Cauchy-Schwarz.

199. DESIGUALDAD DE MINKOWSKI. Demuestre que para cada $r \geq 1$,

$$E^{1/r} |X + Y|^r \leq E^{1/r} |X|^r + E^{1/r} |Y|^r.$$

Sugerencia: $E |X + Y|^r \leq E (|X| \cdot |X + Y|^{r-1}) + E (|Y| \cdot |X + Y|^{r-1})$, ahora use la desigualdad de Hölder.

Cuantiles

200. Calcule los cuantiles de la distribución normal estándar.

201. Calcule los cuantiles de la distribución exponencial de parámetro λ .

202. MINIMIZACIÓN DEL ERROR ABSOLUTO MEDIO. A la función $g(u) = E |X - u|$ se le conoce como error absoluto medio. Demuestre que si m una mediana de X , entonces para cualquier número real u ,

$$E |X - m| \leq E |X - u|.$$

Demuestre además que la igualdad se cumple si, y sólo si, u es cualquier otra mediana de X .

203. Sea X una variable aleatoria con segundo momento finito y sea m una de sus medianas. Demuestre que $|m - E(X)| \leq \sqrt{2 \text{Var}(X)}$.

Distribución uniforme discreta

204. Sea X con distribución $\text{unif}\{1, \dots, n\}$. Demuestre que

a) $E(X) = (n + 1)/2$.

b) $E(X^2) = (n + 1)(2n + 1)/6$.

c) $\text{Var}(X) = (n^2 - 1)/12$.

205. Se escogen al azar y de manera independiente dos números a y b dentro del conjunto $\{1, \dots, n\}$. Demuestre que la probabilidad de que el cociente a/b sea menor o igual a uno es $(n + 1)/2n$.

Distribución Bernoulli

206. Compruebe que la función de probabilidad de la distribución $\text{Ber}(p)$ efectivamente lo es. Obtenga además la correspondiente función de distribución. Grafique ambas funciones.
207. Sea X con distribución $\text{Ber}(p)$. Demuestre que $E(X^n) = p$, para cada $n \geq 1$. En particular, compruebe que $\text{Var}(X) = p(1 - p)$.

Distribución binomial

208. Use el teorema del binomio para comprobar que la función de probabilidad de la distribución $\text{bin}(n, p)$ efectivamente lo es.
209. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- $E(X) = np$.
 - $E(X^2) = np(1 - p + np)$.
 - $\text{Var}(X) = np(1 - p)$.
 - $E(X - np)^3 = np(1 - p)(1 - 2p)$.
 - $E(X - np)^4 = 3n^2p^2(1 - p)^2 + np(1 - p)(1 - 6(1 - p)p)$.
210. Sea X con distribución $\text{bin}(n, p)$. Demuestre que $Y = n - X$ tiene distribución $\text{bin}(n, 1 - p)$.
211. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- $P(X = x + 1) = \frac{p}{1 - p} \frac{n - x}{x + 1} P(X = x)$.
 - $P(X = x - 1)P(X = x + 1) \leq P^2(X = x)$.
212. Sea X con distribución $\text{bin}(n, p)$. Demuestre que
- $P(X \in \{1, 3, 5, \dots\}) = \frac{1}{2}(1 - (1 - 2p)^n)$.

$$b) P(X \in \{0, 2, 4, \dots\}) = \frac{1}{2}(1 + (1 - 2p)^n).$$

213. Se lanza una moneda equilibrada 6 veces. Calcule la probabilidad de que cada cara se obtenga exactamente 3 veces.

Distribución geométrica

214. Compruebe que la función de probabilidad de la distribución $\text{geo}(p)$ efectivamente lo es. Demuestre que la correspondiente función de distribución es

$$F(x) = \begin{cases} 1 - (1 - p)^{\lfloor x \rfloor + 1} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

La expresión $\lfloor x \rfloor$ denota la parte entera de x .

215. Sea X con distribución $\text{geo}(p)$. Demuestre que

$$\begin{aligned} a) E(X) &= (1 - p)/p. \\ b) \text{Var}(X) &= (1 - p)/p^2. \end{aligned}$$

216. Sea X con distribución $\text{geo}(p)$. Demuestre que $P(X \geq n) = (1 - p)^n$. Use este resultado y la fórmula del ejercicio 157 para demostrar que $E(X) = (1 - p)/p$.

217. LA DISTRIBUCIÓN GEOMÉTRICA NO TIENE MEMORIA. Sea X con distribución $\text{geo}(p)$. Demuestre que para cualesquiera $x, y = 0, 1, \dots$

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Esta es la única distribución discreta con tal propiedad, compare con el siguiente ejercicio.

218. Sea X una variable aleatoria discreta con valores en $\{0, 1, \dots\}$ y tal que para cualquier $x, y = 0, 1, \dots$ se cumple la igualdad

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Demuestre que existe un número $p \in (0, 1)$ tal que X tiene distribución $\text{geo}(p)$.

Distribución Poisson

219. Compruebe que la función de probabilidad de la distribución Poisson(λ) efectivamente lo es.
220. Sea X con distribución Poisson(λ). Demuestre que
- $E(X) = \lambda$.
 - $E(X^2) = \lambda(\lambda + 1)$.
 - $\text{Var}(X) = \lambda$.
 - $E(X^3) = \lambda E(X + 1)^2$.
221. Sea X con distribución Poisson(λ). Demuestre que
- $P(X = x + 1) = \frac{\lambda}{x + 1} P(X = x)$.
 - $P(X = x - 1) P(X = x + 1) \leq P^2(X = x)$.
222. Sea X con distribución Poisson(λ). Demuestre que
- $P(X \in \{1, 3, 5, \dots\}) = \frac{1}{2}(1 - e^{-2\lambda})$.
 - $P(X \in \{0, 2, 4, \dots\}) = \frac{1}{2}(1 + e^{-2\lambda})$.
223. TEOREMA DE POISSON (CONVERGENCIA DE LA DIST. BINOMIAL A LA DIST. POISSON). Para cada entero positivo n , sea X_n con distribución $\text{bin}(n, \lambda/n)$ con $\lambda > 0$. Demuestre que para cada $k = 0, 1, \dots$

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

A este resultado también se le conoce con el nombre de ley de eventos raros.

Distribución binomial negativa

224. Compruebe que la función de probabilidad de la distribución bin neg(r, p) efectivamente lo es.
225. Sea X con distribución bin neg(r, p). Demuestre que
- $E(X) = r(1 - p)/p$.
 - $\text{Var}(X) = r(1 - p)/p^2$.
226. CONVERGENCIA DE LA DIST. BINOMIAL NEGATIVA A LA DIST. POISSON. Sea X_1, X_2, \dots una sucesión de variables tal que cada una de ellas tiene distribución bin neg(n, p) con $p = n/(\lambda + n)$ para algún $\lambda > 0$. Demuestre que para cada $k = 0, 1, \dots$

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Distribución hipergeométrica

227. Compruebe que la función de probabilidad de la distribución hipergeométrica efectivamente lo es.
228. CONVERGENCIA DE LA DIST. HIPERGEOMÉTRICA A LA DIST. BINOMIAL. Sea X con distribución hipergeo(N, K, n). Demuestre que cuando N y K tienden a infinito de tal forma que $K/N \rightarrow p$, entonces

$$\lim_{N, K \rightarrow \infty} P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

Distribución uniforme continua

229. Compruebe que la función de densidad de la distribución unif(a, b) efectivamente lo es. Calcule además la correspondiente función de distribución. Grafique ambas funciones.

230. Sea X con distribución $\text{unif}(a, b)$. Demuestre que

a) $E(X) = (a + b)/2$.

b) $E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$.

c) $\text{Var}(X) = (b-a)^2/12$.

231. Sea X con distribución $\text{unif}(0, 1)$. Demuestre que $E(X^n) = 1/(n+1)$.

232. Sea X con distribución $\text{unif}(-1, 1)$. Demuestre que para $n = 0, 1, 2, \dots$

$$E(X^n) = \begin{cases} 1/n + 1 & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

233. Sea X con distribución $\text{unif}(0, 1)$. Obtenga la distribución de

a) $Y = 10X - 5$.

b) $Y = 4X(1 - X)$.

234. Sea X con distribución $\text{unif}(0, 1)$ y sea $0 < p < 1$. Demuestre que la variable aleatoria $Y = \lfloor \ln X / \ln(1 - p) \rfloor$ tiene distribución $\text{geo}(p)$. La expresión $\lfloor x \rfloor$ denota la parte entera de x .

235. Sea X con distribución $\text{unif}(0, 1)$. Defina a Y como el primer dígito decimal de X . Demuestre que Y tiene distribución uniforme en el conjunto $\{0, 1, \dots, 9\}$.

Distribución exponencial

236. Compruebe que la función de densidad de la distribución $\text{exp}(\lambda)$ efectivamente lo es. Demuestre que la correspondiente función de distribución es

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Demuestre además que para cualquier $x, y > 0$,

$$F(x+y) - F(y) = F(x)(1 - F(y)).$$

237. Demuestre que la esperanza de la distribución $\exp(\lambda)$ es $1/\lambda$, y la varianza es $1/\lambda^2$.

238. LA DISTRIBUCIÓN EXPONENCIAL NO TIENE MEMORIA. Sea X con distribución $\exp(\lambda)$. Demuestre que

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

La distribución exponencial es la única distribución absolutamente continua que satisface esta propiedad, al respecto ver el siguiente ejercicio.

239. Sea X una variable aleatoria absolutamente continua con valores en el intervalo $(0, \infty)$, y tal que para cualesquiera $x, y > 0$ se cumple

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Demuestre que existe una constante $\lambda > 0$ tal que X tiene distribución $\exp(\lambda)$.

240. Sea X una variable aleatoria con función de distribución continua $F(x)$, estrictamente creciente y tal que $0 < F(x) < 1$. Demuestre que la variable aleatoria $Y = -\ln F(X)$ tiene distribución exponencial con parámetro $\lambda = 1$.

241. Sea $a > 0$. Demuestre que si X se distribuye $\exp(\lambda)$, entonces aX se distribuye $\exp(\lambda/a)$.

242. Se dice que la variable X tiene una *distribución exponencial bilateral* (o *exponencial doble*) con parámetro $\lambda > 0$ si su función de densidad es

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \quad \text{para } x \in \mathbb{R}.$$

Demuestre que la esperanza de esta distribución es cero, y la varianza es $2/\lambda^2$.

243. Sea X una variable aleatoria con distribución exponencial de parámetro λ , y sea a una constante positiva. Calcule la esperanza y varianza de la variable $\min\{X, a\}$.

Distribución gama

244. Compruebe que la función de densidad de la distribución $\text{gama}(n, \lambda)$ efectivamente lo es. Verifique además que esta distribución se reduce a la distribución $\text{exp}(\lambda)$ cuando $n = 1$.
245. Sea $a > 0$. Demuestre que si X se distribuye $\text{gama}(n, \lambda)$, entonces aX se distribuye $\text{gama}(n, \lambda/a)$.
246. Sea X con distribución $\text{gama}(n, \lambda)$. Demuestre que la función de distribución de X es

$$F(x) = \begin{cases} 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

247. Sea X con distribución $\text{gama}(n, \lambda)$. Demuestre que

- a) $E(X) = n/\lambda$.
 b) $E(X^m) = \frac{\Gamma(m+n)}{\lambda^m \Gamma(n)}$, para $m = 0, 1, \dots$
 c) $\text{Var}(X) = n/\lambda^2$.

248. Recuerde que la función gama se define para cada valor de n tal que la siguiente integral es convergente

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t} dt.$$

Demuestre que esta función cumple las siguientes propiedades.

- a) $\Gamma(n+1) = n\Gamma(n)$.
 b) $\Gamma(n+1) = n!$ para n entero.
 c) $\Gamma(2) = \Gamma(1) = 1$.
 d) $\Gamma(1/2) = \sqrt{\pi}$.
 e) $\Gamma(n+1/2) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \sqrt{\pi}$ para n entero.

Distribución beta

249. Compruebe que la función de densidad de la distribución beta(a, b) efectivamente lo es. Verifique además que esta distribución se reduce a la distribución unif($0, 1$) cuando $a = b = 1$.

250. Sea X con distribución beta(a, b). Demuestre que

$$a) E(X) = \frac{a}{a+b}.$$

$$b) E(X^n) = \frac{B(a+n, b)}{B(a, b)}.$$

$$c) \text{Var}(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

251. Sea X con distribución beta(a, b). Demuestre que

$$a) a = E(X) \left[\frac{E(X)(1-E(X))}{\text{Var}(X)} - 1 \right].$$

$$b) b = (1-E(X)) \left[\frac{E(X)(1-E(X))}{\text{Var}(X)} - 1 \right].$$

$$c) a+b = \frac{E(X)(1-E(X))}{\text{Var}(X)} - 1.$$

252. Recuerde que la función beta se define para cada $a, b > 0$ de la forma

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx.$$

Demuestre que esta función cumple las siguientes propiedades.

$$a) B(a, b) = B(b, a).$$

$$b) B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b).$$

$$c) B(a, 1) = 1/a.$$

$$d) B(1, b) = 1/b.$$

$$e) B(a+1, b) = \frac{a}{b} B(a, b+1).$$

$$f) B(a+1, b) = \frac{a}{a+b} B(a, b).$$

$$g) B(a, b+1) = \frac{b}{a+b} B(a, b).$$

$$h) B(1/2, 1/2) = \pi.$$

253. Sea X con distribución beta(1/2, 1/2). En este caso se dice que X tiene una *distribución arcoseno*.

a) Calcule y grafique $f(x)$.

b) Demuestre directamente que $f(x)$ es una función de densidad.

c) Demuestre directamente que $E(X) = 1/2$, y $\text{Var}(X) = 1/8$.

254. Sea X con distribución beta(a, b). Demuestre que para $a > 0$ y $b = 1$,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^a & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

255. Sea X con distribución beta(a, b). Demuestre que para $a = 1$ y $b > 0$,

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - (1-x)^b & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

256. Demuestre que X tiene distribución beta(a, b) si, y sólo si, $1 - X$ tiene distribución beta(b, a).

Distribución normal

257. Demuestre que la función de densidad de la distribución $N(\mu, \sigma^2)$

a) es efectivamente una función de densidad.

b) es simétrica respecto de $x = \mu$.

c) alcanza su máximo en $x = \mu$.

d) tiene puntos de inflexión en $x = \mu \pm \sigma$.

258. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que $E(X) = \mu$ y $\text{Var}(X) = \sigma^2$.

259. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que para cada $n = 0, 1, 2, \dots$

$$E|X - \mu|^n = \begin{cases} 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

260. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que

a) $P(\mu - \sigma < X < \mu + \sigma) = 0.68269$.

b) $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$.

c) $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$.

261. Sea X con distribución normal estándar. Demuestre que para cada $n = 0, 1, \dots$

$$E(X^n) = \begin{cases} \frac{n!}{2^{n/2}(n/2)!} & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

262. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que $Y = aX + b$, con $a \neq 0$, tiene una distribución normal. Encuentre los parámetros correspondientes.

263. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que la variable aleatoria $-X$ también tiene una distribución normal. Encuentre los parámetros correspondientes.

264. Sea X con distribución normal estándar. Demuestre que X^2 tiene una distribución $\chi^2(1)$. Recíprocamente, ¿Será cierto que si Y tiene distribución, $\chi^2(1)$ entonces \sqrt{Y} tiene distribución $N(0, 1)$?

265. Encuentre la función de densidad de la variable aleatoria $|X|$, cuando X tiene distribución normal estándar.

266. EL COCIENTE DE MILLS. Sea $\phi(x)$ la función de densidad de la distribución normal estándar, y sea $\Phi(x)$ la correspondiente función de distribución. Demuestre que

a) $\phi'(x) + x\phi(x) = 0$.

b) $\frac{1}{x} - \frac{1}{x^3} < \frac{1 - \Phi(x)}{\phi(x)} < \frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}$, para $x > 0$.

Distribución log normal

267. Demuestre que la función de densidad de una distribución log normal (μ, σ^2) efectivamente lo es.

268. Sea X con distribución log normal (μ, σ^2) . Demuestre que

a) $E(X) = \exp(\mu + \sigma^2/2)$.

b) $\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)$.

c) $E(\ln X) = \mu$.

d) $\text{Var}(\ln X) = \sigma^2$.

CAPÍTULO 3

Vectores aleatorios

En este capítulo se extiende el concepto de variable aleatoria con valores reales a variables aleatorias con valores en \mathbb{R}^n , a tales funciones las llamaremos vectores aleatorios. Estudiaremos además varios conceptos importantes relacionados con estas funciones.

3.1. Vectores aleatorios

Recuerde que hemos supuesto que se tiene siempre como elemento base un espacio de probabilidad (Ω, \mathcal{F}, P) .

DEFINICIÓN. (VECTOR ALEATORIO). Un vector aleatorio es una función $X : \Omega \rightarrow \mathbb{R}^n$ tal que para cualquier conjunto B en $\mathcal{B}(\mathbb{R}^n)$, se cumple que la imagen inversa $X^{-1}B$ es un elemento de \mathcal{F} .

Dado entonces que un vector aleatorio es una función de Ω en \mathbb{R}^n , éste puede representar de la forma $X = (X_1, \dots, X_n)$ en donde cada coordenada es una función de Ω en \mathbb{R} . Demostraremos a continuación que la condición que aparece en la definición anterior es equivalente a solicitar que cada

coordenada de este vector sea una variable aleatoria. En consecuencia, es correcto definir un vector aleatorio simplemente como un vector de variables aleatorias. Véase la Figura 3.1. Puede demostrarse además que si se parte de n espacios de probabilidad en donde están definidas n variables aleatorias respectivamente, entonces existe un espacio de probabilidad, el espacio de probabilidad producto, en donde el vector aleatorio está definido.

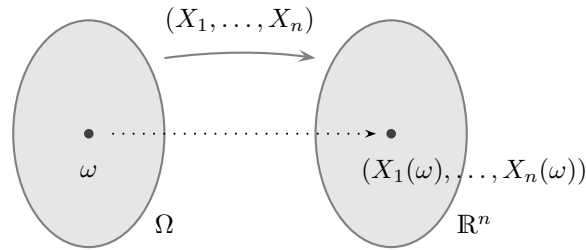


Figura 3.1: Un vector aleatorio es una función de Ω en \mathbb{R}^n .

PROPOSICIÓN. Una función $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ es un vector aleatorio si, y sólo si, cada coordenada es una variable aleatoria.

Demostración. Sea (X_1, \dots, X_n) un vector aleatorio. La imagen inversa de cualquier conjunto de Borel de \mathbb{R}^n es entonces un elemento de la σ -álgebra del espacio de probabilidad. En particular, la imagen inversa del conjunto $B \times \mathbb{R} \times \dots \times \mathbb{R}$ pertenece a \mathcal{F} , para cualquier Boreliano B de \mathbb{R} . Pero esta imagen inversa es simplemente $X_1^{-1}B$. Esto demuestra que X_1 es variable aleatoria. De manera análoga se procede con las otras coordenadas del vector. Suponga ahora que cada coordenada de una función $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ es una variable aleatoria. Considere la colección $\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}^n) : (X_1, \dots, X_n)^{-1}B \in \mathcal{F}\}$. Como cada coordenada es una variable aleatoria, los conjuntos de Borel de \mathbb{R}^n de la forma $B_1 \times \dots \times B_n$, en donde cada factor de este producto es un Boreliano de \mathbb{R} , es un elemento

de la colección \mathcal{B} . Entonces

$$\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R}) \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R}^n).$$

Es fácil demostrar que la colección \mathcal{B} es una σ -álgebra. Así que

$$\sigma(\mathcal{B}(\mathbb{R}) \times \cdots \times \mathcal{B}(\mathbb{R})) \subseteq \mathcal{B} \subseteq \mathcal{B}(\mathbb{R}^n).$$

Pero ambos extremos de esta ecuación coinciden, de modo que $\mathcal{B} = \mathcal{B}(\mathbb{R}^n)$, y por lo tanto la función (X_1, \dots, X_n) es un vector aleatorio. \square

Para simplificar la escritura, donde sea posible se usan únicamente vectores aleatorios bidimensionales, esto es, de la forma (X, Y) . En la mayoría de los casos, las definiciones y resultados son fácilmente extendidos a dimensiones mayores. Por ejemplo, el siguiente resultado es análogo al caso unidimensional y puede extenderse al caso de n dimensiones: un vector aleatorio $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ genera el espacio de probabilidad $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), P_{X,Y})$, en donde $\mathcal{B}(\mathbb{R}^2)$ es la σ -álgebra de conjuntos de Borel de \mathbb{R}^2 , y $P_{X,Y}$ es una medida de probabilidad definida sobre esta σ -álgebra, e inducida por el vector aleatorio de la siguiente forma. Para cualquier B en $\mathcal{B}(\mathbb{R}^2)$,

$$P_{X,Y}(B) = P((X, Y)^{-1}B).$$

Nuestro objetivo es estudiar estas nuevas medidas de probabilidad, o equivalentemente, los vectores aleatorios que las generan. En la mayoría de los casos, aunque no únicamente, consideraremos vectores aleatorios como los que se definen a continuación.

DEFINICIÓN. (VECTOR DISCRETO Y CONTINUO). Se dice que el vector (X, Y) es discreto si cada coordenada es una variable aleatoria discreta, y se dice que es continuo en caso de que cada coordenada lo sea.

3.2. Distribución conjunta

Como en el caso de variables aleatorias, todo vector aleatorio induce una medida de probabilidad, ahora sobre \mathbb{R}^n . Esta medida de probabilidad puede estudiarse, de manera equivalente, mediante la función de distribución conjunta definida a continuación.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN CONJUNTA). La función de distribución de un vector (X, Y) , denotada por $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$, se define como sigue

$$F(x, y) = P(X \leq x, Y \leq y).$$

El número $F(x, y)$ es entonces la probabilidad de que el vector aleatorio tome algún valor en la rectángulo infinito $(-\infty, x] \times (-\infty, y]$, el cual se muestra en la Figura 3.2. En palabras, la función $F(x, y)$ es la probabilidad de que X sea menor o igual a x , y al mismo tiempo Y sea menor o igual a y , esto es simplemente la probabilidad del evento $(X \leq x) \cap (Y \leq y)$.

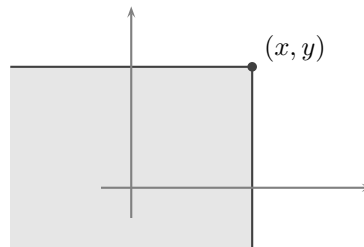


Figura 3.2: El número $F(x, y) = P(X \leq x, Y \leq y)$ es la probabilidad de que el vector (X, Y) tome un valor en la región sombreada.

A la función $F(x, y)$ se le conoce también como función de distribución *bivariada* de X y Y , y en general a la distribución conjunta de un vector aleatorio

de cualquier dimensión finita se le llama distribución *multivariada*. Naturalmente, en el caso unidimensional, la distribución se llama *univariada*. Cuando sea necesario especificarlo se escribe $F_{X,Y}(x,y)$ en lugar de $F(x,y)$, y es evidente la forma de extender la definición para el caso de vectores aleatorios de más de dos coordenadas. Con el fin de mantener la notación simple, en la medida de lo posible se mantiene la correspondencia de las letras, es decir, x es un valor asociado a X , y y es un valor asociado a Y .

Las funciones de distribución conjunta satisfacen propiedades semejantes al caso unidimensional, se estudian a continuación algunas de ellas.

PROPOSICIÓN. Toda función de distribución conjunta $F(x,y)$ satisface las siguientes propiedades.

1. $\lim_{x,y \rightarrow \infty} F(x,y) = 1$, ambas variables.
2. $\lim_{x,y \rightarrow -\infty} F(x,y) = 0$, alguna de las variables.
3. $F(x,y)$ es no decreciente en cada variable.
4. $F(x,y)$ es continua por la derecha en cada variable.
5. Si $a_1 < b_1$ y $a_2 < b_2$, entonces

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0.$$

La demostración de las propiedades (1) a (4) es completamente análoga al caso unidimensional y por tanto la omitiremos. Respecto a la propiedad (5) observe que la expresión $F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$ corresponde a la probabilidad del evento $(a_1 < X \leq b_1, a_2 < Y \leq b_2)$. De modo que (5) se traduce simplemente en solicitar que la probabilidad de que el vector (X, Y) tome valores en el rectángulo $(a_1, b_1] \times (a_2, b_2]$, sea no negativa. Este rectángulo se muestra en la Figura 3.3.

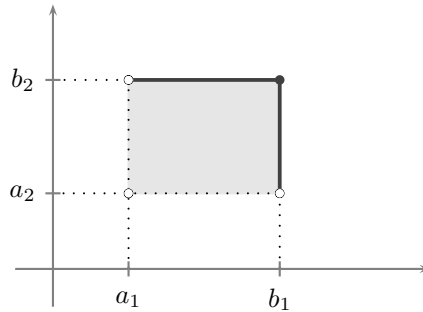


Figura 3.3: La probabilidad asociada al rectángulo $(a_1, b_1] \times (a_2, b_2]$ es $P(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$.

EJERCICIO. Grafique y demuestre que la siguiente función es de distribución.

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{si } x, y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

▪

A diferencia del caso unidimensional, las propiedades (1) a (4) no son suficientes para asegurar que una función $F(x, y)$ asigna probabilidad no negativa a cualquier rectángulo. El siguiente ejercicio muestra un ejemplo de esta situación. Véase también el ejercicio 272.

EJERCICIO. Grafique y demuestre que la función que aparece abajo no es de distribución. Este es un ejemplo de una función que tiene el comportamiento límite adecuado en infinito, es continua por la derecha y no decreciente en cada variable, pero no es función de distribución pues asigna valores negativos a algunas regiones del plano. Por ejemplo calcule la probabilidad del cuadrado $(-1, 1] \times (-1, 1]$.

$$F(x, y) = \begin{cases} 0 & \text{si } x + y < 0, \\ 1 & \text{si } x + y \geq 0. \end{cases}$$

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN CONJUNTA). Una función cualquiera $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$, no necesariamente definida en términos de un vector aleatorio, es una función de distribución conjunta si cumple con las cinco propiedades enunciadas en la proposición anterior.

Más adelante se mostrarán otros ejemplos concretos de funciones de distribución conjunta.

Para tres dimensiones se tiene la siguiente definición. Se dice que la función $F : \mathbb{R}^3 \rightarrow [0, 1]$ es una *función de distribución* si cumple las primeras cuatro propiedades anteriores y la quinta propiedad se reemplaza por la siguiente condición: Para cualesquiera números reales $a_1 < b_1$, $a_2 < b_2$, y $a_3 < b_3$,

$$\begin{aligned} &F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ &+ F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) \\ &- F(a_1, a_2, a_3) \geq 0. \end{aligned}$$

En términos de vectores aleatorios se puede demostrar que el lado izquierdo de esta desigualdad corresponde a la probabilidad del evento $(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3)$, es decir, se trata de la probabilidad de que el vector aleatorio (X_1, X_2, X_3) tome algún valor dentro del paralelepípedo que se muestra en la Figura 3.4. La condición anterior establece entonces que este número debe ser mayor o igual a cero.

Más generalmente, se tiene la siguiente definición.

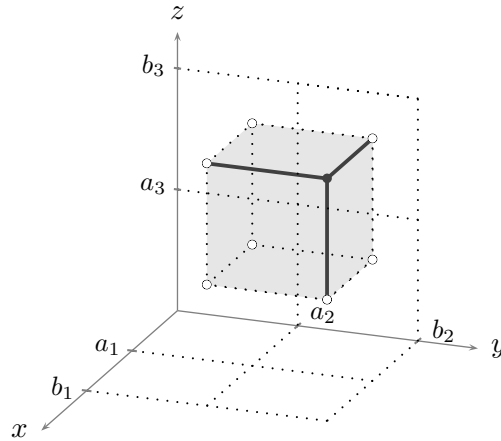


Figura 3.4: Región $(a_1, b_1] \times (a_2, b_2] \times (a_3, b_3]$.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN CONJUNTA). Una función $F : \mathbb{R}^n \rightarrow [0, 1]$ es una función de distribución si cumple las primeras cuatro propiedades anteriores y, adicionalmente, para cualesquiera números reales $a_1 < b_1$, $a_2 < b_2$, \dots , $a_n < b_n$,

$$\sum_{x_i \in \{a_i, b_i\}} (-1)^{\#a} F(x_1, \dots, x_n) \geq 0,$$

en donde $\#a$ es el número de veces que alguna de las variables x_i toma el valor a_i en la evaluación de la función F .

Nuevamente la suma que aparece en esta definición corresponde a la probabilidad del evento $(a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n)$, y la condición requiere simplemente que este número sea no negativo.

Finalmente enunciamos un resultado que establece la importancia de la función de distribución, y cuya demostración puede ser encontrada por ejemplo en [19]. La prueba no es sencilla pero es análoga al caso unidimensional.

PROPOSICIÓN. Sea $F : \mathbb{R}^n \rightarrow [0, 1]$ una función de distribución. Entonces existe un espacio de probabilidad, y un vector aleatorio, cuya función de distribución es F .

Es decir, este resultado garantiza la existencia de un espacio de probabilidad (Ω, \mathcal{F}, P) en donde se encuentra definido un vector aleatorio (X_1, \dots, X_n) con función de distribución la especificada. En lo que resta del capítulo hablaremos de vectores aleatorios suponiendo que existe un espacio de probabilidad base asociado.

3.3. Densidad conjunta

Como en el caso unidimensional, algunos vectores tienen asociada otra función llamada de probabilidad o de densidad, y la cual se define a continuación.

DEFINICIÓN. (FUNCIÓN DE PROBABILIDAD CONJUNTA). La función de probabilidad de un vector discreto (X, Y) es la función $f(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$ dada por

$$f(x, y) = P(X = x, Y = y).$$

A esta función también se le llama función de probabilidad conjunta de las variables X y Y .

Es evidente que la función de probabilidad de un vector discreto cumple las siguientes propiedades.

- a) $f(x, y) \geq 0$.
- b) $\sum_{x, y} f(x, y) = 1$.

Recíprocamente, toda función no negativa $f(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$ que sea es-

trictamente positiva únicamente en un subconjunto discreto de \mathbb{R}^2 y que sume uno, se llama *función de probabilidad conjunta*. La definición de función de probabilidad en el caso discreto multidimensional es evidente. Es claro también que la correspondiente función de distribución se puede calcular a partir de la función de probabilidad de la siguiente forma:

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v).$$

EJEMPLO. La función $f(x, y) = 1/4$, para $x, y = 1, 2$, es una función de probabilidad conjunta pues es no negativa y suma uno, corresponde a la distribución uniforme sobre el conjunto $\{1, 2\} \times \{1, 2\}$. La gráfica se muestra en la Figura 3.5.

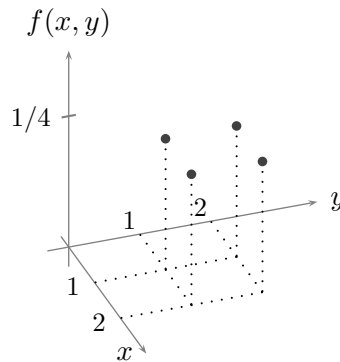


Figura 3.5: Función de probabilidad $f(x, y) = 1/4$, para $x, y = 1, 2$.

La correspondiente función de distribución es

$$F(x, y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v) = \begin{cases} 0 & \text{si } x < 1 \text{ ó } y < 1, \\ 1/4 & \text{si } 1 \leq x < 2, 1 \leq y < 2, \\ 2/4 & \text{si } 1 \leq x < 2, y \geq 2, \\ 2/4 & \text{si } x \geq 2, 1 \leq y < 2, \\ 1 & \text{si } x \geq 2 \text{ y } y \geq 2, \end{cases}$$

cuya gráfica se encuentra en la Figura 3.6.

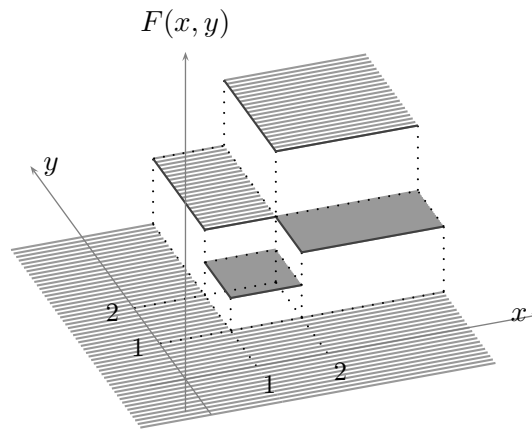


Figura 3.6: Ejemplo de función de distribución discreta.

EJEMPLO. La función definida por $f(x, y) = (1/2)^{x+y}$ para $x, y \in \mathbb{N}$, e idénticamente cero fuera de este conjunto discreto, es una función de probabilidad bivariada pues es no negativa y suma uno. En efecto,

$$\sum_{x,y=1}^{\infty} f(x, y) = \sum_{x,y=1}^{\infty} \frac{1}{2^{x+y}} = \left(\sum_{x=1}^{\infty} \frac{1}{2^x} \right)^2 = 1.$$

Para el caso de vectores continuos se tiene la siguiente definición.

DEFINICIÓN. (FUNCIÓN DE DENSIDAD CONJUNTA). Sea (X, Y) un vector continuo con función de distribución $F(x, y)$. Se dice que (X, Y) es absolutamente continuo si existe una función no negativa e integrable $f(x, y) : \mathbb{R}^2 \rightarrow [0, \infty)$, tal que, para todo (x, y) en \mathbb{R}^2 , se cumple la igualdad

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du.$$

A la función $f(x, y)$ se le denota por $f_{X,Y}(x, y)$, y se le llama función de densidad conjunta de X y Y .

Así como en el caso unidimensional, no existe realmente unicidad para la función de densidad pues basta modificarla en algunos puntos para ser distinta pero seguir cumpliendo la igualdad anterior, sin embargo la función de distribución y por tanto las probabilidades, permanecen sin cambio alguno. Es claro que la función de densidad conjunta $f(x, y)$ de un vector absolutamente continuo cumple las siguientes propiedades.

- a) $f(x, y) \geq 0$.
- b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Recíprocamente, toda función no negativa $f : \mathbb{R}^2 \rightarrow [0, \infty)$, que integre uno, se llama *función de densidad conjunta*. En particular, cuando $f(x, y)$ es continua,

$$f(x, y) = \frac{\partial^2}{\partial y \partial x} F(x, y).$$

Observe que, en el caso absolutamente continuo y conociendo la función de densidad conjunta, la probabilidad del evento $(a \leq X \leq b, c \leq Y \leq d)$ no cambia si se incluyen o se excluyen los extremos de cada intervalo, y se calcula como la integral doble que se ilustra en la Figura 3.7.

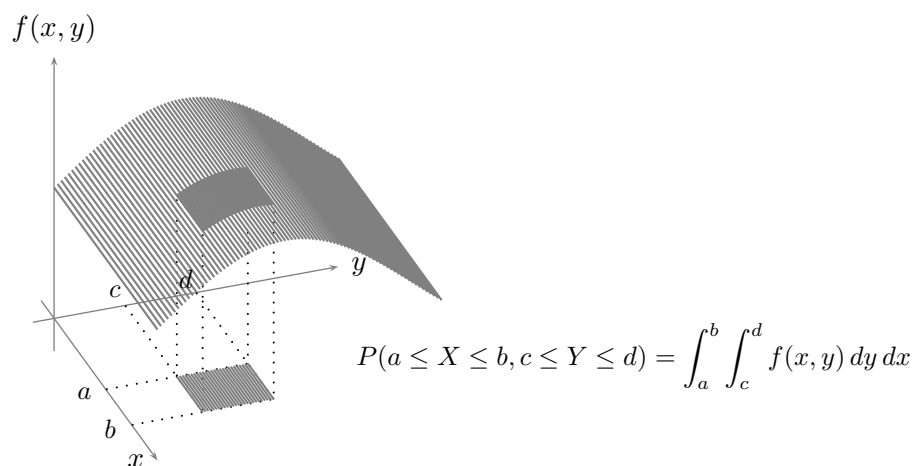


Figura 3.7: La probabilidad como el volumen bajo una superficie.

EJEMPLO. La función $f : \mathbb{R}^2 \rightarrow [0, \infty)$ dada por la siguiente expresión es una función de densidad pues es no negativa e integra uno.

$$f(x, y) = \begin{cases} 1/4 & \text{si } x, y \in [0, 2], \\ 0 & \text{otro caso.} \end{cases}$$

Esta función de densidad conjunta corresponde a la distribución uniforme del vector (X, Y) en el cuadrado $[0, 2] \times [0, 2]$. La gráfica se muestra en la Figura 3.8.

Para calcular la correspondiente función de distribución $F(x, y)$ se debe calcular la doble integral para los distintos valores de x y y . Esta doble integral toma distintas expresiones en cada una de las cinco regiones que aparecen en la parte izquierda de la Figura 3.9. Después de algunos cálculos elementales se encuentra que la función de distribución conjunta tiene la

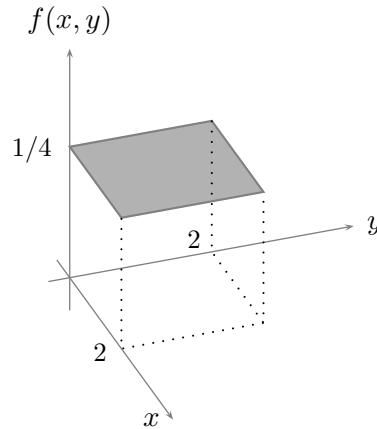


Figura 3.8: Función de densidad $f(x, y) = 1/4$, para $x, y \in [0, 2]$.

siguiente expresión cuya gráfica aparece en la parte derecha de la Figura 3.9.

$$\begin{aligned}
 F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du \\
 &= \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0, \\ xy/4 & \text{si } 0 \leq x, y < 2, \\ x/2 & \text{si } 0 \leq x < 2, y \geq 2, \\ y/2 & \text{si } 0 \leq y < 2, x \geq 2, \\ 1 & \text{si } x \geq 2 \text{ y } y \geq 2. \end{cases}
 \end{aligned}$$

▪

EJERCICIO. Demuestre que la siguiente función es de densidad. Calcule $P(X = Y)$, $P(X = 1)$ y $P(X + Y = n + 1)$. ¿Puede usted calcular la

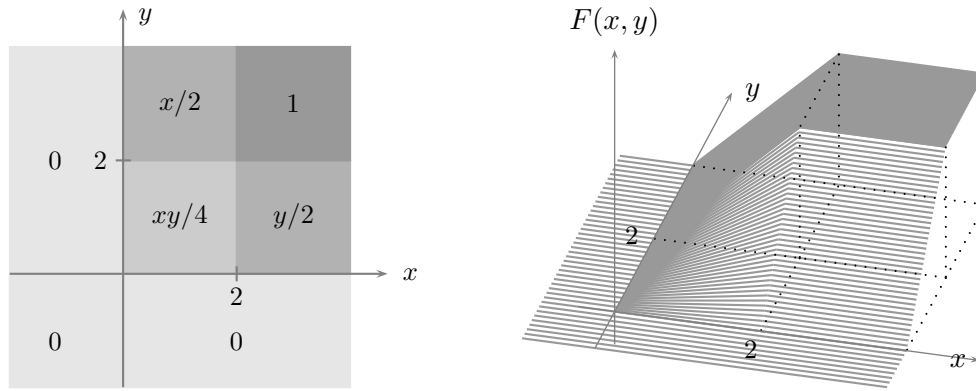


Figura 3.9: Ejemplo de función de distribución continua bivariada.

correspondiente función de distribución?

$$f(x, y) = \begin{cases} \frac{4xy}{n^2(n+1)^2} & \text{si } x, y = 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}$$

EJERCICIO. Demuestre que la siguiente función es de densidad. Encuentre la correspondiente función de distribución y grafique ambas funciones. Calcule además $P(1/3 < X < 1, 0 < Y < 1/2)$, $P(Y > X)$ y $P(X > 1/2)$.

$$f(x, y) = \begin{cases} x + y & \text{si } 0 < x, y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

EJERCICIO. Demuestre que la siguiente función es de densidad. Encuentre la correspondiente función de distribución y calcule $P(1 < X < 2, 1 < Y < 2)$, $P(X > 1)$ y $P(Y + X > 2)$.

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

3.4. Distribución marginal

Dada la función de distribución $F(x, y)$ de un vector aleatorio (X, Y) , es posible obtener la función de distribución de cada variable aleatoria por separado mediante el siguiente procedimiento.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN MARGINAL). Sea (X, Y) un vector con función de distribución $F(x, y)$. A la función

$$F(x) = \lim_{y \rightarrow \infty} F(x, y)$$

se le conoce como la función de distribución marginal de X . Análogamente se define la función de distribución marginal de Y como

$$F(y) = \lim_{x \rightarrow \infty} F(x, y).$$

No es difícil verificar que las funciones de distribución marginales son efectivamente funciones de distribución univariadas. En el caso de vectores de dimensión mayor, se puede obtener la distribución marginal de cualquier subconjunto de variables aleatorias del vector original mediante un procedimiento similar.

EJEMPLO. En un ejemplo anterior habíamos encontrado la siguiente función

de distribución conjunta

$$F_{X,Y}(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0, \\ xy/4 & \text{si } 0 \leq x, y < 2, \\ x/2 & \text{si } 0 \leq x < 2, y \geq 2, \\ y/2 & \text{si } 0 \leq y < 2, x \geq 2, \\ 1 & \text{si } x \geq 2 \text{ y } y \geq 2. \end{cases}$$

Esta función está definida de manera distinta en cada una de las cinco regiones disjuntas y exhaustivas del plano Cartesiano dadas por las condiciones anteriores. Para encontrar, por ejemplo, la función de distribución marginal $F_X(x)$ simplemente tenemos que hacer la variable y tender a infinito en las regiones donde ello sea posible. Ello puede hacerse en las regiones dadas por las condiciones del primer, tercer y quinto renglón de la lista anterior. Esto da como resultado

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ x/2 & \text{si } 0 \leq x < 2, \\ 1 & \text{si } x \geq 2. \end{cases}$$

¿Puede usted encontrar ahora $F_Y(y)$? ▪

EJERCICIO. Encuentre las funciones de distribución marginales del vector (X, Y) cuya función de distribución es

$$F(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0, \\ 3x^2y/5 + 2xy^3/5 & \text{si } 0 \leq x < 1 \text{ y } 0 \leq y < 1, \\ 3x^2/5 + 2x/5 & \text{si } 0 \leq x < 1 \text{ y } y \geq 1, \\ 3y/5 + 2y^3/5 & \text{si } x \geq 1 \text{ y } 0 \leq y < 1, \\ 1 & \text{si } x \geq 1 \text{ y } y \geq 1. \end{cases}$$

Para el caso de funciones de densidad conjunta, se pueden obtener las funciones de densidad individuales como indica la siguiente definición.

DEFINICIÓN. (FUNCIÓN DE DENSIDAD MARGINAL). Sea (X, Y) un vector absolutamente continuo con función de densidad $f(x, y)$. A la función

$$f(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

se le conoce como la función de densidad marginal de X . Análogamente se define la función de densidad marginal de Y como

$$f(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Si (X, Y) es un vector discreto la integral se reemplaza por una suma.

Tampoco es difícil comprobar que las funciones de densidad marginales son efectivamente funciones de densidad univariadas. Las dos definiciones anteriores pueden extenderse de manera evidente cuando se tenga un vector aleatorio de cualquier dimensión finita. También es posible calcular las funciones de densidad y de distribución de (X, Y) a partir, por ejemplo, de las funciones correspondientes del vector (X, Y, Z) .

EJERCICIO. Calcule las funciones de densidad marginales del vector aleatorio discreto (X, Y) cuya función de probabilidad esta dada por la siguiente tabla.

$x \backslash y$	1	2	3
-1	1/45	2/45	3/45
0	4/45	5/45	6/45
1	7/45	8/45	9/45

▪

EJERCICIO. Calcule las funciones de densidad marginales del vector aleatorio continuo (X, Y) cuya función de densidad es

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

Observe que la distribución conjunta determina de manera única a las distribuciones marginales. Sin embargo, si lo que se conoce son las distribuciones marginales, entonces puede haber varias distribuciones conjuntas que produzcan las marginales dadas. La forma de producir la distribución conjunta se llama *acoplamiento*, y la distribución conjunta obtenida se llama a veces *distribución de acoplamiento* o *cópula*. Dos variables aleatorias X y Y siempre pueden acoplarse de la forma $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, que es el caso donde se han hecho independientes una de la otra, pero puede haber otras formas de hacerlo. En el siguiente ejemplo se muestra una situación concreta en el caso discreto.

EJEMPLO. Sean X y Y discretas ambas con distribución uniforme en el conjunto $\{0, 1\}$, es decir, su distribución de probabilidad es

$$f(x) = \begin{cases} 1/2 & \text{si } x = 0, 1, \\ 0 & \text{otro caso.} \end{cases}$$

Sean $a \geq 0$ y $b \geq 0$ tales que $a + b = 1/2$. Entonces la siguiente densidad conjunta tiene como densidades marginales las especificadas para X y para Y .

$x \backslash y$	0	1
0	a	b
1	b	a

Observe que esta densidad conjunta es en realidad toda una familia de densidades conjuntas que producen las densidades marginales especificadas. En este caso X y Y son independientes si, y sólo si, $a = b = 1/4$.

3.5. Distribución condicional

La siguiente definición es una extensión del concepto elemental de probabilidad condicional de eventos.

DEFINICIÓN. (FUNCIÓN DE DENSIDAD CONDICIONAL). Sea (X, Y) un vector con función de densidad $f_{X,Y}(x, y)$, y sea y tal que $f_Y(y) \neq 0$. A la función

$$x \mapsto f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

se le conoce como la función de densidad condicional de X dado que Y toma el valor y .

No es difícil comprobar que esta función es efectivamente una función de densidad, tanto en el caso discreto como en el continuo. Observe que el valor y permanece fijo y la función es vista como una función de la variable real x , esto puede observarse en el siguiente ejemplo.

EJEMPLO. Considere la función de densidad conjunta

$$f_{X,Y}(x, y) = \begin{cases} 24x(1-y) & \text{si } 0 < x < y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

Es sencillo comprobar que para cualquier valor fijo de y en el intervalo $(0, 1)$, la función de densidad condicional de X dado Y es la que aparece más abajo. Es también inmediato verificar que esta función, vista como función de x , es de densidad. El valor de y puede entonces considerarse como un parámetro de esta nueva distribución.

$$f_{X|Y}(x|y) = \begin{cases} 2x/y^2 & \text{si } 0 < x < y, \\ 0 & \text{otro caso.} \end{cases}$$

Análogamente puede comprobarse que para cualquier x en $(0, 1)$ fijo,

$$f_{Y|X}(y|x) = \begin{cases} 2(1-y)/(x-1)^2 & \text{si } x < y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

EJERCICIO. Calcule las funciones de densidad condicionales $f_{Y|X}(y|x)$ y $f_{X|Y}(x|y)$ a partir de la siguiente función de densidad conjunta

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

Se pueden definir también funciones de distribución condicionales de la siguiente forma.

DEFINICIÓN. (FUNCIÓN DE DISTRIBUCIÓN CONDICIONAL). Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad $f_{X,Y}(x, y)$, y sea y tal que $f_Y(y) \neq 0$. A la función

$$x \mapsto F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du$$

se le conoce como la función de distribución condicional de X dado que Y toma el valor y . Cuando el vector aleatorio (X, Y) es discreto la integral se substituye por la suma correspondiente.

Nuevamente resulta que la función de distribución condicional es efectivamente una función de distribución. En el caso absolutamente continuo y suponiendo $x \mapsto f_{X|Y}(x|y)$ continua, por el teorema fundamental del cálculo se tiene que

$$f_{X|Y}(x|y) = \frac{\partial}{\partial x} F_{X|Y}(x|y).$$

EJEMPLO. Considere nuevamente la función de densidad conjunta del ejemplo anterior, $f_{X,Y}(x, y) = 24x(1 - y)$, para $0 < x < y < 1$. Para y fijo en el

intervalo $(0, 1)$ se tiene que

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du = \begin{cases} 0 & \text{si } x \leq 0, \\ x^2/y^2 & \text{si } 0 < x < y, \\ 1 & \text{si } x \geq y. \end{cases}$$

Puede también definirse la *esperanza condicional* de la siguiente forma.

DEFINICIÓN. Sea (X, Y) un vector con función de distribución $F_{X,Y}(x, y)$, y sea y un valor tal que $f_Y(y) \neq 0$. Si X tiene esperanza finita, entonces se define

$$E(X | Y = y) = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y).$$

En el siguiente capítulo veremos una definición mucho más general de este concepto.

EJERCICIO. Calcule la función de distribución condicional $F_{X|Y}(x|y)$ a partir de la función de densidad conjunta $f_{X,Y}(x, y) = 3(x^2 + y^2)/16$, para $0 < x < y < 2$. Calcule además $E(X | Y = y)$ para cualquier valor de y en el intervalo $(0, 2)$.

EJERCICIO. Calcule $E(X | Y = y)$ para $y = \pi/4$, cuando (X, Y) es un vector absolutamente continuo con función de densidad $f(x, y) = (1/2) \sin(x + y)$, para $x, y \in (0, \pi/2)$.

EJERCICIO. Sea (X, Y) un vector aleatorio tal que X tiene esperanza finita y Y es discreta con valores $0, 1, \dots$ tal que $P(Y = n) > 0$ para $n = 0, 1, \dots$. Demuestre que

$$E(X) = \sum_{n=0}^{\infty} E(X | Y = n) P(Y = n).$$

3.6. Independencia

Podemos ahora definir el importante concepto de independencia de variables aleatorias. Primero definiremos tal concepto para dos variables aleatorias, después lo haremos para n variables, y finalmente para una colección arbitraria de variables aleatorias.

DEFINICIÓN. (INDEPENDENCIA DE DOS VARIABLES ALEATORIAS). Se dice que X y Y son independientes, y a menudo se escribe $X \perp Y$, si para cada par de conjuntos de Borel A, B de \mathbb{R} , se cumple la igualdad

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B). \quad (3.1)$$

En términos de la siempre existente función de distribución, la independencia de dos variables aleatorias se puede expresar como indica el siguiente resultado.

PROPOSICIÓN. (INDEPENDENCIA DE DOS VARIABLES ALEATORIAS). Las variables aleatorias X y Y son independientes si, y sólo si, para cada (x, y) en \mathbb{R}^2 se cumple la igualdad

$$F_{X,Y}(x, y) = F_X(x) F_Y(y). \quad (3.2)$$

Demostración. Si X y Y son independientes, entonces tomando $A = (-\infty, x]$ y $B = (-\infty, y]$ en (3.1) se obtiene (3.2). Suponga ahora que se cumple (3.2)

para cualesquiera x y y en \mathbb{R} . Defina la colección

$$\mathcal{A} = \{A \in \mathcal{B}(\mathbb{R}) : P(X \in A, Y \leq y) = P(X \in A) P(Y \leq y), \forall y \in \mathbb{R}\}.$$

No es difícil demostrar que \mathcal{A} es una σ -álgebra y usando la hipótesis resulta que $\mathcal{A} = \mathcal{B}(\mathbb{R})$. Sea ahora A un elemento cualquiera fijo de $\mathcal{B}(\mathbb{R})$. Defina la colección

$$\mathcal{B} = \{B \in \mathcal{B}(\mathbb{R}) : P(X \in A, Y \in B) = P(X \in A) P(Y \in B)\}.$$

Se puede comprobar nuevamente que \mathcal{B} es una σ -álgebra, y de hecho $\mathcal{B} = \mathcal{B}(\mathbb{R})$. De esta forma, para cualquier A y B en $\mathcal{B}(\mathbb{R})$, se cumple la condición (3.1). \square

El concepto de independencia de variables aleatorias es una extensión de la misma propiedad para eventos. Cuando la función de densidad conjunta existe, la condición de independencia de X y Y es equivalente a solicitar que para cualesquiera números reales x y y , se cumpla la identidad

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \quad (3.3)$$

En el caso discreto, la afirmación anterior es completamente correcta. Para el caso continuo hay una observación técnica que es necesario mencionar. Como en este caso las funciones de densidad pueden ser modificadas sin que cambie la función de distribución asociada, la igualdad (3.3) puede no cumplirse para cada $(x, y) \in \mathbb{R}^2$, entonces se permite que la igualdad no se cumpla en un conjunto de medida de Lebesgue cero, por ejemplo, un conjunto numerable de parejas (x, y) en \mathbb{R}^2 , y entonces habrá independencia en el caso continuo si se cumple (3.3), salvo conjuntos de medida de Lebesgue cero.

EJEMPLO. Sea (X, Y) un vector aleatorio con función de densidad $f(x, y) = 4xy$, para $0 \leq x, y \leq 1$, y cuya gráfica aparece en la Figura 3.10.

La función de densidad marginal de X se calcula de la siguiente forma. Para $0 \leq x \leq 1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 4xy dy = 2x.$$

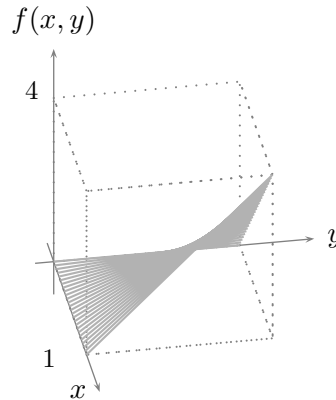


Figura 3.10: Función de densidad $f(x, y) = 4xy$, para $0 \leq x, y \leq 1$.

Análogamente $f_Y(y) = 2y$ para $0 \leq y \leq 1$. En consecuencia, X y Y son independientes pues para cada par (x, y) , se cumple $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

EJERCICIO. Determine si las variables aleatorias continuas X y Y son independientes cuando su función de densidad conjunta es

$$f_{X,Y}(x, y) = \begin{cases} 3(x^2 + y^2)/32 & \text{si } 0 < x, y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

El concepto de independencia puede ser extendido claramente al caso de varias variables aleatorias de la forma siguiente.

DEFINICIÓN. (INDEPENDENCIA DE VARIAS VARIABLES ALEATORIAS). Se dice que las variables X_1, \dots, X_n son independientes si para cualesquiera Borelianos A_1, \dots, A_n de \mathbb{R} , se cumple

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Más aún, una colección infinita de variables aleatorias es independiente si cualquier subconjunto finito de ella lo es.

Usando un procedimiento similar al caso de dos variables aleatorias, puede demostrarse que la condición de independencia de n variables aleatorias es equivalente a solicitar que para cualquier vector (x_1, \dots, x_n) en \mathbb{R}^n se cumpla la igualdad

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Y en términos de la función de densidad, cuando ésta exista y salvo un conjunto de medida cero, la condición es

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

Cuando las variables X_1, \dots, X_n son independientes y tomando conjuntos Borelianos adecuados en la definición general, puede comprobarse que cualquier subconjunto de estas variables también son independientes. El recíproco, sin embargo, es en general falso como se pide demostrar a continuación.

EJERCICIO. Sean X y Y independientes ambas con distribución uniforme en el conjunto $\{-1, 1\}$. Sea $Z = XY$. Demuestre que X, Y y Z son independientes dos a dos pero no lo son en su conjunto. ▪

PROPOSICIÓN. Sean X y Y independientes, y sean g y h dos funciones de \mathbb{R} en \mathbb{R} , Borel medibles. Entonces las variables aleatorias $g(X)$ y $h(Y)$ también son independientes.

Demostración. Sean A y B cualesquiera dos conjuntos de Borel de \mathbb{R} . Entonces

$$\begin{aligned} P(g(X) \in A, h(Y) \in B) &= P(X \in g^{-1}(A), Y \in h^{-1}(B)) \\ &= P(X \in g^{-1}(A)) P(Y \in h^{-1}(B)) \\ &= P(g(X) \in A) P(h(Y) \in B). \end{aligned}$$

□

Este resultado puede extenderse fácilmente al caso n -dimensional, y de esta forma obtener que la composición de n funciones Borel medibles aplicadas, respectivamente, a n variables aleatorias independientes, produce nuevamente variables aleatorias independientes.

La definición de independencia de dos variables aleatorias puede extenderse al caso de dos vectores aleatorios de cualquier dimensión de la forma siguiente.

DEFINICIÓN. (INDEPENDENCIA DE DOS VECTORES ALEATORIOS). Se dice que los vectores $X = (X_1, \dots, X_n)$ y $Y = (Y_1, \dots, Y_m)$ son independientes, si para cada A en $\mathcal{B}(\mathbb{R}^n)$, y cada B en $\mathcal{B}(\mathbb{R}^m)$, se cumple la igualdad

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B). \quad (3.4)$$

Naturalmente esta definición puede extenderse un poco más para incluir la independencia de un número finito de vectores aleatorios, no necesariamente todos de la misma dimensión. Y nuevamente, una colección infinita de vectores aleatorios es independiente si cualquier subcolección finita de ellos lo es.

EJERCICIO. Demuestre que si los vectores (X_1, \dots, X_n) y (Y_1, \dots, Y_m) son independientes, entonces las variables X_i y Y_j son independientes para cualquier posible valor de los índices i y j . ■

3.7. Esperanza de una función de un vector aleatorio

Si (X, Y) es un vector aleatorio y $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ es una función Borel medible, entonces $\varphi(X, Y)$ es una variable aleatoria y el problema nuevamente es encontrar su esperanza. Usando directamente la definición, la esperanza de $\varphi(X, Y)$ se calcula del siguiente modo:

$$E[\varphi(X, Y)] = \int_{-\infty}^{\infty} \varphi(x) dF_{\varphi(X, Y)}(x),$$

pero, así como en el caso unidimensional, ello requiere encontrar primero la distribución de $\varphi(X, Y)$, lo cual puede ser difícil en muchos casos. El siguiente resultado establece una forma alternativa de calcular la esperanza de $\varphi(X, Y)$, sin conocer su distribución, pero conociendo, por supuesto, la distribución del vector (X, Y) .

TEOREMA (ESPERANZA DE UNA FUNCIÓN DE UN VECTOR ALEATORIO). Sea (X, Y) un vector aleatorio, y sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función Borel medible tal que la variable aleatoria $\varphi(X, Y)$ tiene esperanza finita. Entonces

$$E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) dF_{X, Y}(x, y). \quad (3.5)$$

Nuevamente omitiremos la demostración de este resultado. Observe que se trata de una integral de Riemann-Stieltjes en dos dimensiones. El “incremento” de F en el rectángulo $(x_{i-1}, x_i] \times (y_{j-1}, y_j]$ es

$$F(x_i, y_j) - F(x_i, y_{j-1}) - F(x_{i-1}, y_j) + F(x_{i-1}, y_{j-1}).$$

Véase nuevamente la Figura 3.3 para comprobar esta expresión. En el caso

cuando X y Y son independientes, este incremento es

$$\begin{aligned} & F(x_i)F(y_j) - F(x_i)F(y_{j-1}) - F(x_{i-1})F(y_j) + F(x_{i-1})F(y_{j-1}) \\ &= (F(x_i) - F(x_{i-1}))(F(y_j) - F(y_{j-1})) \\ &= \Delta F(x_i) \Delta F(y_j), \end{aligned}$$

es decir, la integral bidimensional se separa en dos integrales, y se puede escribir

$$E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) dF_X(x) dF_Y(y).$$

Cuando el vector (X, Y) es discreto, la fórmula (3.5) se reduce a

$$E[\varphi(X, Y)] = \sum_{x, y} \varphi(x, y) P(X = x, Y = y),$$

en donde la suma se efectúa sobre todos los posibles valores (x, y) del vector. En este caso la demostración del teorema resulta no muy complicada, y se pide dar los detalles en el siguiente ejercicio.

EJERCICIO. Sea (X, Y) un vector aleatorio discreto con valores en el conjunto producto $\{x_1, x_2, \dots\} \times \{y_1, y_2, \dots\}$, y sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ una función Borel medible tal que la variable $\varphi(X, Y)$ tiene esperanza finita. Demuestre que

$$E[\varphi(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \varphi(x_i, y_j) P(X = x_i, Y = y_j).$$

▪

En el caso cuando (X, Y) es absolutamente continuo, la expresión (3.5) se escribe

$$E[\varphi(X, Y)] = \int_{\mathbb{R}^2} \varphi(x, y) f_{X, Y}(x, y) dx dy.$$

Con ayuda de este resultado podemos ahora demostrar que la esperanza separa sumas.

PROPOSICIÓN. Sean X y Y con esperanza finita. Entonces

$$E(X + Y) = E(X) + E(Y).$$

Demostración. Sean $\varphi(x, y) = x + y$, $\varphi_1(x, y) = x$, y $\varphi_2(x, y) = y$. Entonces

$$\begin{aligned} E(X + Y) &= E(\varphi(X, Y)) \\ &= \int_{\mathbb{R}^2} (x + y) dF_{X,Y}(x, y) \\ &= \int_{\mathbb{R}^2} x dF_{X,Y}(x, y) + \int_{\mathbb{R}^2} y dF_{X,Y}(x, y) \\ &= E(\varphi_1(X, Y)) + E(\varphi_2(X, Y)) \\ &= E(X) + E(Y). \end{aligned}$$

□

PROPOSICIÓN. Sean X y Y independientes, y sean g y h dos funciones Borel medibles tales que $g(X)$ y $h(Y)$ tienen esperanza finita. Entonces

$$E[g(X)h(Y)] = E[g(X)] E[h(Y)].$$

En particular, $E(XY) = E(X) E(Y)$.

Demostración.

$$\begin{aligned} E[g(X)h(Y)] &= \int_{\mathbb{R}^2} g(x)h(y) dF_{X,Y}(x, y) \\ &= \int_{\mathbb{R}^2} g(x)h(y) dF_X(x) dF_Y(y) \\ &= E[g(X)] E[h(Y)]. \end{aligned}$$

□

En general, el recíproco de la afirmación anterior es falso, es decir, la condición $E(XY) = E(X)E(Y)$ no es suficiente para poder concluir que X y Y son independientes. Por ejemplo, considere el vector aleatorio discreto (X, Y) con función de probabilidad

$x \backslash y$	-1	0	1
-1	1/5	0	1/5
0	0	1/5	0
1	1/5	0	1/5

Es sencillo verificar que $E(XY) = E(X)E(Y) = 0$, sin embargo X y Y no son independientes pues $P(X = 0, Y = 0) = 1/5$, mientras que $P(X = 0)P(Y = 0) = 1/25$. Otros ejemplos de esta misma situación pueden encontrarse en el ejercicio 352 en la página 203.

3.8. Covarianza

En esta sección se define y estudia la covarianza entre dos variables aleatorias. Una interpretación de este número, ligeramente modificado, será dada en la siguiente sección.

DEFINICIÓN. (COVARIANZA). La covarianza de X y Y , denotada por $\text{Cov}(X, Y)$, es el número

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Para que la definición anterior tenga sentido es necesario suponer que las esperanzas $E(X)$, $E(Y)$ y $E(XY)$ son finitas. En general cuando se escribe $\text{Cov}(X, Y)$, se suponen tales condiciones. Se revisan a continuación algunas propiedades de la covarianza.

PROPOSICIÓN. Sean X y Y variables aleatorias y sea c una constante. Entonces

1. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X, X) = \text{Var}(X)$.
4. $\text{Cov}(c, Y) = 0$.
5. $\text{Cov}(cX, Y) = c \text{Cov}(X, Y)$.
6. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.
7. Si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$.
8. En general, $\text{Cov}(X, Y) = 0 \not\Rightarrow X, Y$ independientes.

Demostración.

1. Por la propiedad de linealidad de la esperanza,

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - YE(X) - XE(Y) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

2. - 4. Estas propiedades se siguen directamente de la definición.
5. - 6. Esto es consecuencia de la definición y de la linealidad de la esperanza.
7. Esta propiedad se obtiene fácilmente de la primera pues $E(XY) = E(X)E(Y)$ cuando X y Y son independientes.

8. Sea (X, Y) un vector aleatorio discreto con función de densidad

$$f_{X,Y}(x, y) = \begin{cases} 1/8 & \text{si } (x, y) \in \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}, \\ 1/2 & \text{si } (x, y) = (0, 0), \\ 0 & \text{otro caso.} \end{cases}$$

Entonces X y Y tienen idénticas densidades marginales,

$$f_X(x) = \begin{cases} 1/4 & \text{si } x \in \{-1, 1\}, \\ 1/2 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases} \quad f_Y(y) = \begin{cases} 1/4 & \text{si } y \in \{-1, 1\}, \\ 1/2 & \text{si } y = 0, \\ 0 & \text{otro caso.} \end{cases}$$

Puede entonces comprobarse que $\text{Cov}(X, Y) = 0$. Sin embargo X y Y no son independientes pues en particular $P(X = 0, Y = 0) = 1/2$, mientras que $P(X = 0)P(Y = 0) = 1/4$.

□

Observe en particular que la covarianza es una función bilineal y simétrica. Estas propiedades serán de utilidad en la siguiente sección. Más adelante demostraremos que, en el caso especial cuando el vector (X, Y) tiene distribución normal bivariada, la condición de covarianza cero implica que estas variables son efectivamente independientes.

EJERCICIO. Sean X y Y independientes. Demuestre que $\text{Cov}(X + Y, Y) = \text{Var}(Y)$. ■

3.9. Coeficiente de correlación

El coeficiente de correlación de dos variables aleatorias es un número real que mide el grado de *dependencia lineal* que existe entre ellas. Su definición es la siguiente.

DEFINICIÓN. (COEFICIENTE DE CORRELACIÓN). El coeficiente de correlación de las variables aleatorias X y Y , denotado por $\rho(X, Y)$, es el número

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Naturalmente en esta definición se necesita suponer que las varianzas son estrictamente positivas y finitas. Vista como función de dos variables aleatorias, el coeficiente de correlación es una función simétrica pero no es lineal pues no separa sumas ni multiplicaciones por constantes.

EJERCICIO. Demuestre que, en general,

- a) $\rho(cX, Y) \neq c\rho(X, Y)$, c constante.
- b) $\rho(X_1 + X_2, Y) \neq \rho(X_1, Y) + \rho(X_2, Y)$.

La interpretación dada al coeficiente de correlación se justifica a partir de los siguientes resultados.

PROPOSICIÓN. El coeficiente de correlación satisface las siguientes propiedades.

1. Si X y Y son independientes, entonces $\rho(X, Y) = 0$.
2. $-1 \leq \rho(X, Y) \leq 1$.
3. $|\rho(X, Y)| = 1$ si, y sólo si, existen constantes a y b tales que, con probabilidad uno, $Y = aX + b$, con $a > 0$ si $\rho(X, Y) = 1$, y $a < 0$ si $\rho(X, Y) = -1$.

Demostración.

1. Si X y Y son independientes, entonces $\text{Cov}(X, Y) = 0$, y por lo tanto $\rho(X, Y) = 0$.
2. Suponga primero que X y Y son tales que $E(X) = E(Y) = 0$, y $\text{Var}(X) = \text{Var}(Y) = 1$. Para cualquier valor de λ ,

$$\begin{aligned} 0 &\leq \text{Var}(X + \lambda Y) \\ &= E(X + \lambda Y)^2 - E^2(X + \lambda Y) \\ &= 1 + 2\lambda E(XY) + \lambda^2. \end{aligned}$$

El caso $\lambda = 1$ produce el resultado $E(XY) \geq -1$, mientras que para $\lambda = -1$ se obtiene $E(XY) \leq 1$. Es decir, $-1 \leq E(XY) \leq 1$. Observe que estas desigualdades también pueden ser obtenidas a partir de la desigualdad de Cauchy-Schwarz. Ahora se aplica este resultado a las variables aleatorias $(X - \mu_X)/\sigma_X$ y $(Y - \mu_Y)/\sigma_Y$, que evidentemente son centradas y con varianza unitaria. Entonces

$$-1 \leq E\left(\frac{X - \mu_X}{\sigma_X} \frac{Y - \mu_Y}{\sigma_Y}\right) \leq 1.$$

El término de enmedio es $\rho(X, Y)$.

3. Si X y Y son tales que $Y = aX + b$ con $a \neq 0$ y b constantes, entonces

$$\rho(X, Y) = \frac{\text{Cov}(X, aX + b)}{\sqrt{\text{Var}(X)\text{Var}(aX + b)}} = \frac{a}{|a|}.$$

Por lo tanto $\rho(X, Y) = 1$ cuando $a > 0$, y $\rho(X, Y) = -1$ cuando $a < 0$. Inversamente, suponga que X y Y son tales que $|\rho(X, Y)| = 1$. Defina $U = (X - \mu_X)/\sigma_X$ y $V = (Y - \mu_Y)/\sigma_Y$. Entonces claramente $E(U) = E(V) = 0$, y $\text{Var}(U) = \text{Var}(V) = 1$. Por lo tanto $\rho(U, V) = E(UV)$. Es fácil ver también que $|\rho(U, V)| = |\rho(X, Y)| = 1$. Si $\rho(U, V) = 1$,

entonces

$$\begin{aligned}\text{Var}(U - V) &= E(U - V)^2 - E^2(U - V) \\ &= E(U - V)^2 \\ &= 2(1 - E(UV)) \\ &= 0.\end{aligned}$$

Esto significa que con probabilidad uno, la variable $U - V$ es constante. Esto es, para alguna constante c , con probabilidad uno, $U - V = c$. Pero esta constante c debe ser cero pues $E(U - V) = 0$. Por lo tanto,

$$\frac{X - \mu_X}{\sigma_X} = \frac{Y - \mu_Y}{\sigma_Y},$$

de donde se obtiene $Y = \mu_Y + (X - \mu_X)\sigma_Y/\sigma_X$. Esto establece una relación lineal directa entre X y Y . En cambio, si $\rho(U, V) = -1$, entonces

$$\begin{aligned}\text{Var}(U + V) &= E(U + V)^2 - E^2(U + V) \\ &= E(U + V)^2 \\ &= 2(1 + E(UV)) \\ &= 0.\end{aligned}$$

Esto significa nuevamente que con probabilidad uno, la variable $U + V$ es constante. Esto es, para alguna constante c , con probabilidad uno, $U + V = c$. Nuevamente la constante c es cero pues $E(U + V) = 0$. Por lo tanto,

$$\frac{X - \mu_X}{\sigma_X} = -\frac{Y - \mu_Y}{\sigma_Y},$$

de donde se obtiene $Y = \mu_Y - (X - \mu_X)\sigma_Y/\sigma_X$. Esto establece una relación lineal, ahora inversa, entre X y Y . Uniendo los últimos dos resultados se obtiene que, cuando $|\rho(X, Y)| = 1$, con probabilidad uno,

$$Y = \left[\rho(X, Y) \frac{\sigma_Y}{\sigma_X} \right] X + \left[\mu_Y - \rho(X, Y) \mu_X \frac{\sigma_Y}{\sigma_X} \right].$$

□

EJERCICIO. Sean X y Y independientes e idénticamente distribuidas. Demuestre que $\rho(X + Y, X - Y) = 0$. ■

DEFINICIÓN. (CORRELACIÓN POSITIVA, NEGATIVA O NULA). Cuando $\rho(X, Y) = 0$ se dice que X y Y son no correlacionadas. Cuando $|\rho(X, Y)| = 1$ se dice que X y Y están perfectamente correlacionadas positiva o negativamente, de acuerdo al signo de $\rho(X, Y)$.

Nuevamente observe que, en general, la condición $\rho(X, Y) = 0$ no es suficiente para poder afirmar que X y Y son independientes, excepto en el caso normal. Esto es consecuencia del mismo resultado para la covarianza.

EJERCICIO. Sea X una variable aleatoria discreta con distribución uniforme en el conjunto $\{-2, -1, 1, 2\}$, y defina $Y = X^2$. Demuestre que el coeficiente de correlación entre X y Y es cero, y sin embargo X y Y no son independientes. ■

Adicionalmente en los ejercicios 380 y 381 de la página 208 se muestran situaciones concretas de este mismo resultado tanto en el caso discreto como en el continuo. Sin embargo, cuando la distribución de (X, Y) es normal y $\rho(X, Y) = 0$, entonces efectivamente se cumple que X y Y son independientes. Demostraremos esto a continuación.

PROPOSICIÓN. Si (X, Y) es un vector con distribución normal bivariada tal que $\rho(X, Y) = 0$, entonces X y Y son independientes.

Demostración. Como veremos más adelante, la función de densidad normal

bivariada está dada por la siguiente expresión:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right]\right),$$

en donde $\mu_1 = E(X)$, $\sigma_1^2 = \text{Var}(X)$, $\mu_2 = E(Y)$, $\sigma_2^2 = \text{Var}(Y)$, y $\rho \in (-1, 1)$. Se pueden calcular directamente las funciones de densidad marginales y comprobar que

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp[-(x-\mu_1)^2/2\sigma_1^2] \\ \text{y} \quad f(y) &= \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp[-(y-\mu_2)^2/2\sigma_2^2], \end{aligned}$$

es decir, X tiene distribución $N(\mu_1, \sigma_1^2)$, y Y tiene distribución $N(\mu_2, \sigma_2^2)$. Después de hacer algunos cálculos sencillos se puede demostrar que el coeficiente de correlación entre X y Y es ρ , y comprobar finalmente que cuando este número es cero, se verifica la igualdad $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, para cualesquiera valores reales de x y y .

□

En resumen tenemos la siguiente tabla.

PROPIEDADES DEL COEFICIENTE DE CORRELACIÓN

$\rho(X, Y) \in [-1, 1]$.

$|\rho(X, Y)| = 1$ si, y sólo si, $Y = aX + b$, con probabilidad uno.

Si $X \perp Y$, entonces $\rho(X, Y) = 0$.

En general, $\rho(X, Y) = 0 \not\Rightarrow X \perp Y$.

Si (X, Y) tiene dist. normal y $\rho(X, Y) = 0$, entonces $X \perp Y$.

3.10. Esperanza y varianza de un vector aleatorio

Los conceptos de esperanza y varianza para una variable aleatoria pueden extenderse al caso de vectores aleatorios de cualquier dimensión de la siguiente forma.

DEFINICIÓN. (ESPERANZA Y VARIANZA DE UN VECTOR). Sea X el vector aleatorio (X_1, \dots, X_n) . Cuando cada coordenada del vector tiene esperanza finita se define la esperanza de X como el vector numérico

$$E(X) = (E(X_1), \dots, E(X_n)).$$

Si cada coordenada del vector aleatorio tiene segundo momento finito, entonces la varianza de X se define como la matriz cuadrada

$$\text{Var}(X) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Var}(X_n) \end{pmatrix}_{n \times n}.$$

La varianza de un vector X puede expresarse como sigue

$$E [(X - E(X))^t (X - E(X))],$$

en donde X^t significa transpuesta del vector renglón X . Observe que $(X - E(X))^t$ es un vector columna de dimensión $n \times 1$, mientras que $(X - E(X))$ es un vector renglón de dimensión $1 \times n$. De modo que el producto de estos dos vectores, en el orden indicado, resulta en una matriz cuadrada de dimensión $n \times n$ cuya entrada (i, j) es $E[(X_i - E(X_i))(X_j - E(X_j))] = \text{Cov}(X_i, X_j)$. Esta matriz también se llama *matriz de varianzas y covarianzas*, y tiene las siguientes propiedades.

PROPOSICIÓN. La matriz $\text{Var}(X)$ es simétrica y positiva definida. Esto último significa que para cualquier vector $\theta = (\theta_1, \dots, \theta_n)$ de \mathbb{R}^n se cumple la desigualdad $\langle \text{Var}(X)\theta, \theta \rangle \geq 0$, en donde $\langle \cdot, \cdot \rangle$ denota el producto interior usual de \mathbb{R}^n .

Demostración. La simetría se sigue de la igualdad $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. La propiedad de ser positiva definida se obtiene usando la bilinealidad de la covarianza,

$$\begin{aligned} \langle \text{Var}(X)\theta, \theta \rangle &= \sum_{i,j=1}^n \text{Cov}(X_i, X_j)\theta_i\theta_j \\ &= \sum_{i,j=1}^n \text{Cov}(\theta_i X_i, \theta_j X_j) \\ &= \text{Cov}\left(\sum_{i=1}^n \theta_i X_i, \sum_{j=1}^n \theta_j X_j\right) \\ &= \text{Var}\left(\sum_{i=1}^n \theta_i X_i\right) \geq 0. \end{aligned}$$

□

Se puede también definir la *matriz de coeficientes de correlación* del vector X como sigue

$$\rho(X) = \begin{pmatrix} \rho(X_1, X_1) & \cdots & \rho(X_1, X_n) \\ \vdots & & \vdots \\ \rho(X_n, X_1) & \cdots & \rho(X_n, X_n) \end{pmatrix}_{n \times n}$$

A esta matriz también se le llama matriz de correlación. Cumple la propiedad de ser simétrica y tener todos los elementos de la diagonal iguales a uno.

3.11. Distribuciones multivariadas discretas

En esta sección se estudian algunas distribuciones discretas de vectores aleatorios. Estas distribuciones son ejemplos particulares de medidas de probabilidad sobre \mathbb{R}^n , para algún valor natural de n .

DISTRIBUCIÓN MULTINOMIAL. Suponga que se tiene un experimento aleatorio con k posibles resultados distintos. Las probabilidades para cada uno de estos resultados son respectivamente p_1, \dots, p_k , en donde $p_1 + \cdots + p_k = 1$. Ahora suponga que se tienen n ensayos sucesivos independientes del experimento anterior, y defina las variables aleatorias discretas X_1, \dots, X_k , como aquellas que registran el número de veces que se obtienen cada uno de los k posibles resultados en los n ensayos. Observe que la última variable X_k está determinada por las anteriores pues $X_k = n - X_1 - \cdots - X_{k-1}$. Entonces se dice que el vector $X = (X_1, \dots, X_{k-1})$ tiene una distribución multinomial(n, p_1, \dots, p_{k-1}), y su función de densidad es

$$f(x_1, \dots, x_{k-1}) = \begin{cases} \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} & \text{si } x_1, \dots, x_k = 0, 1, \dots, n \\ & \text{con } x_1 + \cdots + x_k = n, \\ 0 & \text{otro caso.} \end{cases}$$

Los parámetros de esta distribución son entonces el número de ensayos n , y las $k - 1$ probabilidades p_1, \dots, p_{k-1} . El factor que aparece en paréntesis en la función de densidad conjunta se conoce como *coeficiente multinomial* y se define como sigue

$$\binom{n}{x_1 \cdots x_k} = \frac{n!}{x_1! \cdots x_k!}.$$

En particular, se dice que el vector (X_1, X_2) tiene *distribución trinomial* con parámetros (n, p_1, p_2) si su función de densidad es

$$f(x_1, x_2) = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

para $x_1, x_2 = 0, 1, \dots, n$, tales que $x_1 + x_2 \leq n$.

En el caso general no es difícil comprobar que la distribución marginal de la variable X_i es $\text{bin}(n, p_i)$, para $i = 1, \dots, k - 1$. Puede además demostrarse que

$$E(X) = (np_1, \dots, np_{k-1}),$$

$$\text{y } [\text{Var}(X)]_{ij} = \begin{cases} np_i(1 - p_i) & \text{si } i = j, \\ -np_i p_j & \text{si } i \neq j. \end{cases}$$

Observe que cuando únicamente hay dos posibles resultados en cada ensayo, es decir $k = 2$, la distribución multinomial se reduce a la distribución binomial.

DISTRIBUCIÓN HIPERGEOMÉTRICA MULTIVARIADA. Suponga que se tienen N objetos de los cuales N_1 son de un primer tipo, N_2 son de un segundo tipo y así sucesivamente con N_k objetos de tipo k . Entonces $N_1 + \dots + N_k = N$. Suponga que de la totalidad de objetos se obtiene una muestra sin reemplazo de tamaño n , y defina la variables X_1, \dots, X_k , como aquellas que representan el número de objetos seleccionados de cada tipo. Se dice entonces que el vector $X = (X_1, \dots, X_k)$ tiene una distribución hipergeométrica

multivariada y su función de densidad es

$$f(x_1, \dots, x_k) = \frac{\binom{N_1}{x_1} \cdots \binom{N_k}{x_k}}{\binom{N}{n}},$$

en donde cada variable x_i toma valores en el conjunto $\{0, 1, \dots, n\}$ pero sujeto a la condición $x_i \leq N_i$, y en donde además debe cumplirse que $x_1 + \cdots + x_k = n$. Se dice entonces que el vector (X_1, \dots, X_k) tiene distribución hipergeométrica multivariada (N, N_1, \dots, N_k, n) . Observe que cuando únicamente hay dos tipos de objetos, es decir $k = 2$, la distribución hipergeométrica multivariada se reduce a la distribución hipergeométrica univariada. En la sección de ejercicios aparecen expresiones para la esperanza y varianza de esta distribución.

3.12. Distribuciones multivariadas continuas

Ahora estudiamos algunas distribuciones continuas de vectores aleatorios.

DISTRIBUCIÓN UNIFORME BIVARIADA. Se dice que las variables aleatorias continuas X y Y tienen una distribución conjunta uniforme en el rectángulo $(a, b) \times (c, d)$, si su función de densidad es

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{si } x \in (a, b), y \in (c, d), \\ 0 & \text{otro caso.} \end{cases}$$

Se escribe $(X, Y) \sim \text{unif}(a, b) \times (c, d)$. Se puede observar inmediatamente que las distribuciones marginales son nuevamente uniformes, además X y Y siempre son independientes. Es fácil también comprobar que $E(X, Y) = ((a+b)/2, (c+d)/2)$, y que

$$\text{Var}(X, Y) = \begin{pmatrix} (b-a)^2/12 & 0 \\ 0 & (d-c)^2/12 \end{pmatrix}.$$

De manera evidente esta distribución puede extenderse al caso de n dimensiones conservándose las mismas propiedades mencionadas.

DISTRIBUCIÓN NORMAL BIVARIADA. Se dice que las variables aleatorias continuas X y Y tienen una distribución normal bivariada si su función de densidad conjunta es

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right]\right),$$

para cualesquiera valores reales de x y y , y en donde $-1 < \rho < 1$, $\sigma_1 > 0$, $\sigma_2 > 0$, y μ_1, μ_2 son dos constantes reales sin restricción. Se escribe entonces $(X, Y) \sim N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Cuando $\mu_1 = \mu_2 = 0$, y $\sigma_1 = \sigma_2 = 1$, la distribución se llama normal bivariada *estándar*, y su gráfica se muestra en la Figura 3.11 cuando $\rho = 0$. En el siguiente ejercicio se enuncian algunas propiedades de esta distribución.

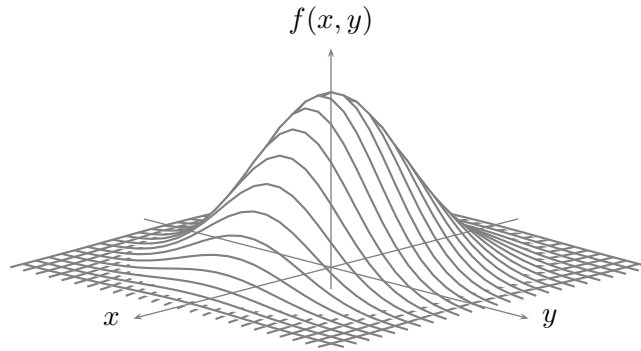


Figura 3.11: Función de densidad normal bivariada estándar.

EJERCICIO. Sea (X, Y) un vector con distribución $N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Demuestre que

- a) X tiene distribución marginal $N(\mu_1, \sigma_1^2)$.
- b) Y tiene distribución marginal $N(\mu_2, \sigma_2^2)$.
- c) $\rho(X, Y) = \rho$.
- d) X y Y son independientes si, y sólo si, $\rho = 0$.
- e) $E(X, Y) = (\mu_1, \mu_2)$.
- f) $\text{Var}(X, Y) = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$.

.

Es interesante observar que existen distribuciones bivariadas con densidades marginales normales, pero cuya distribución conjunta no lo es. En el ejercicio 398 en la página 211 se presenta un ejemplo al respecto.

DISTRIBUCIÓN NORMAL MULTIVARIADA. Se dice que el vector (X_1, \dots, X_n) tiene una distribución normal multivariada si su función de densidad es

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} \exp \left[-\frac{1}{2} (\underline{x} - \underline{\mu}) \Sigma^{-1} (\underline{x} - \underline{\mu})^t \right],$$

en donde $\underline{x} = (x_1, \dots, x_n)$ y $\underline{\mu} = (\mu_1, \dots, \mu_n)$ son dos vectores de números reales, Σ es una matriz de dimensión $n \times n$ positiva definida, es decir, $\underline{x} \Sigma \underline{x}^t \geq 0$ para cualquier vector $\underline{x} = (x_1, \dots, x_n)$ de \mathbb{R}^n , y Σ^{-1} es la matriz inversa de Σ . Como es usual, \underline{x}^t denota el vector transpuesto del vector renglón \underline{x} . Cuando $n = 1$ o $n = 2$, con Σ adecuada, se obtienen las distribuciones normal univariada y bivariada mencionadas antes.

3.13. Ejercicios

Vectores aleatorios

269. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y sea $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ una función tal que cada coordenada es una variable aleatoria. Demuestre que la siguiente colección es una sub σ -álgebra de $\mathcal{B}(\mathbb{R}^n)$.

$$\{B \in \mathcal{B}(\mathbb{R}^n) : (X_1, \dots, X_n)^{-1}B \in \mathcal{F}\}.$$

Distribución conjunta

270. Grafique y demuestre que las siguientes funciones son de distribución.

a) $F(x, y) = (1 - e^{-x})\left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y\right)$, para $x > 0, y \in \mathbb{R}$.

b) $F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}$, para $x, y > 0$.

271. Investigue si las siguientes funciones son de distribución.

a) $F(x, y) = 1 - e^{-xy}$, para $x, y > 0$.

b) $F(x, y) = 1 - e^{-x-y}$, para $x, y > 0$.

272. Demuestre que la siguiente función no es de distribución. Extienda este resultado al caso n -dimensional.

$$F(x, y, z) = \begin{cases} 0 & \text{si } x + y + z < 0, \\ 1 & \text{si } x + y + z \geq 0. \end{cases}$$

273. Demuestre que la siguiente función no es de distribución.

$$F(x, y) = \begin{cases} \min\{1, \max\{x, y\}\} & \text{si } x, y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

274. Sean $F(x)$ y $G(x)$ dos funciones de distribución. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones.

- a) $F(x)G(x)$ es una función de distribución univariada.
- b) $F(x)G(y)$ es una función de distribución bivariada.
- c) $F^n(x)$ es una función de distribución univariada.
- d) $F^n(x)G^m(y)$ es una función de distribución bivariada.

275. Sean X y Y dos variables aleatorias, y sean x y y cualesquiera números reales. Diga falso o verdadero. Demuestre en cada caso.

- a) $P(X > x, Y > y) = 1 - P(X \leq x, Y \leq y)$.
- b) $P(X \leq x, Y \leq y) \leq P(X \leq x)$.
- c) $P(X \leq x) = P(X \leq x, Y \leq x) + P(X \leq x, Y > x)$.
- d) $P(X + Y \leq x) \leq P(X \leq x)$.
- e) $P(XY < 0) \leq P(X < 0)$.

276. Sean X y Y variables aleatorias con función de distribución conjunta $F(x, y)$. Demuestre que para cualesquiera números reales $a < b$ y $c < d$,

$$P(a < X \leq b, c < Y \leq d) = F(b, d) + F(a, c) - F(a, d) - F(b, c).$$

277. Sean X_1, X_2 y X_3 variables aleatorias con función de distribución conjunta $F(x_1, x_2, x_3)$. Demuestre que para cualesquiera números reales $a_1 < b_1, a_2 < b_2$ y $a_3 < b_3$,

$$\begin{aligned} & P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, a_3 < X_3 \leq b_3) \\ &= F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) \\ & \quad + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) \\ & \quad - F(a_1, a_2, a_3). \end{aligned}$$

278. Demuestre que para cualquier vector aleatorio (X, Y) y para cualquier valor real de u , $F_{X,Y}(u, u) = F_{X \vee Y}(u)$, es decir, la función de distribución del máximo de dos variables aleatorias es la distribución conjunta evaluada en la diagonal. Extienda este resultado al caso n -dimensional.

279. Sea (X, Y) un vector con función de distribución $F(x, y)$, y con distribuciones marginales $F(x)$ y $F(y)$, respectivamente. Demuestre que para todo x y y en \mathbb{R} ,

$$F(x) + F(y) - 1 \leq F(x, y) \leq \sqrt{F(x)F(y)}.$$

280. COTAS DE FRÉCHET. Sea (X, Y) un vector con función de distribución $F(x, y)$, y con distribuciones marginales $F(x)$ y $F(y)$, respectivamente. Demuestre que para todo x y y en \mathbb{R} ,

$$\max\{F(x) + F(y) - 1, 0\} \leq F(x, y) \leq \min\{F(x), F(y)\}.$$

281. Considere el espacio $\Omega = (0, 1) \times (0, 1)$ junto con la σ -álgebra $\mathcal{B}((0, 1) \times (0, 1))$ y P la medida de probabilidad uniforme sobre Ω . Sea (X, Y) el vector aleatorio definido sobre este espacio de probabilidad dado por $X(\omega_1, \omega_2) = \omega_1 \wedge \omega_2$ y $Y(\omega_1, \omega_2) = \omega_1 \vee \omega_2$. Demuestre que (X, Y) es efectivamente un vector aleatorio y encuentre su función de distribución.

Densidad conjunta

282. Demuestre que la función de densidad de un vector (X, Y) absolutamente continuo puede ser encontrada, a partir de la función de distribución, de las siguientes formas alternativas:

$$a) f(x, y) = \frac{\partial^2}{\partial x \partial y} P(X > x, Y > y).$$

$$b) f(x, y) = -\frac{\partial^2}{\partial x \partial y} P(X \leq x, Y > y).$$

$$c) f(x, y) = -\frac{\partial^2}{\partial x \partial y} P(X > x, Y \leq y).$$

283. Grafique y demuestre que las siguientes funciones son de densidad.

$$a) f(x, y) = \frac{1}{ab}, \quad \text{para } 0 < x < a, \quad 0 < y < b.$$

- b) $f(x, y) = 4xy$, para $0 \leq x, y \leq 1$.
- c) $f(x, y) = 6x^2y$, para $0 \leq x, y \leq 1$.
- d) $f(x, y) = 9x^2y^2/4$, para $-1 \leq x, y \leq 1$.
- e) $f(x, y) = e^{-x-y}$, para $x, y > 0$.
- f) $f(x, y) = e^{-x}$, para $0 < y < x$.

284. Calcule la constante c que hace a f una función de densidad.

- a) $f(x) = cx$, para $0 \leq x \leq 1$.
- b) $f(x, y) = cx$, para $0 < y < x < 1$.
- c) $f(x, y) = c(x + y)$, para $0 \leq x, y \leq 1$.
- d) $f(x, y) = c(1 - x)(1 - y)$ para $-1 < x, y < 1$.
- e) $f(x, y) = cx(y - x)$ para $0 < x < y < 1$.
- f) $f(x, y) = c(x^2 + \frac{1}{2}xy)$, para $0 < x < 1$, $0 < y < 2$.
- g) $f(x, y, z) = c(x + y + z)$, para $0 \leq x, y, z \leq 1$.
- h) $f(x, y, z) = cx(y - x)(z - y)$, para $0 < x < y < z < 1$.
- i) $f(x_1, \dots, x_n) = c(x_1 + \dots + x_n)$, para $0 \leq x_1, \dots, x_n \leq 1$.

285. Encuentre la función de densidad del vector (X, Y) cuya función de distribución es

- a) $F(x, y) = (1 - e^{-x})(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y)$, para $x > 0$, $y \in \mathbb{R}$.
- b) $F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}$, para $x, y > 0$.

286. Encuentre la función de distribución del vector (X, Y) cuya función de densidad es

- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
- b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.
- c) $f(x, y) = e^{-y}$, para $0 < x < y$.
- d) $f(x, y) = 2e^{-x-y}$, para $0 < x < y$.

287. Sean $f(x)$ y $g(x)$ dos funciones de densidad. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- a) $x \mapsto f(x)g(x)$ es una función de densidad univariada.
- b) $(x, y) \mapsto f(x)g(y)$ es una función de densidad bivariada.

288. Sea (X, Y, Z) un vector con función de densidad

$$f(x, y, z) = \begin{cases} \lambda^3 e^{-\lambda(x+y+z)} & \text{si } x, y, z > 0, \\ 0 & \text{otro caso,} \end{cases}$$

en donde λ es una constante. Calcule $P(X < 1, Y > 2, Z < 3)$, $P(X > 2, Y < 1)$ y $P(X + Y + Z < 1)$.

289. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Encuentre la función de densidad y de distribución de las variables $X \wedge Y$ y $X \vee Y$, cada una de ellas por separado y después de manera conjunta.

Distribución marginal

290. Suponiendo el caso absolutamente continuo, demuestre que la función de densidad marginal $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$ es efectivamente una función de densidad univariada.

291. Demuestre que la función de distribución marginal

$$x \mapsto F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$$

es efectivamente una función de distribución univariada.

292. Encuentre las funciones de distribución marginales del vector (X, Y) cuya función de distribución es

- a) $F(x, y) = (1 - e^{-x})(1 - e^{-y})$, para $x, y > 0$.
- b) $F(x, y) = (1 - e^{-x^2})(1 - e^{-y^2})$, para $x, y > 0$.

293. Encuentre las funciones de densidad marginales del vector (X, Y) cuya función de densidad es

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = 4xy$, para $0 < x, y < 1$.

c) $f(x, y) = 24x(1 - x - y)$, para $x, y > 0$ y $x + y < 1$.

d) $f(x, y) = (x + 2y)/4$, para $0 < x < 2$ y $0 < y < 1$.

e) $f(x, y) = 2(4x + y)/5$, para $0 < x, y < 1$.

f) $f(x, y) = 1/x$, para $0 < y < x < 1$.

g) $f(x, y) = 3/2$, para $0 < y < x^2 < 1$.

h) $f(x, y) = 2x/y^2$, para $0 < x < 1$ y $y > 1$.

294. Encuentre la constante c que hace a f una función de densidad. Encuentre además las funciones de densidad marginales, la función de distribución conjunta asociada y las funciones de distribución marginales.

a) $f(x, y) = c \min\{x, y\}$ para $0 < x, y < 1$.

b) $f(x, y) = c \max\{x + y - 1, 0\}$ para $0 < x, y < 1$.

295. Sea $0 < a < 1$ y defina la función $f(x, y) = a^x(1 - a)^y$, para $x, y = 1, 2, \dots$. Demuestre que $f(x, y)$ es una función de densidad y calcule las funciones de densidad y de distribución marginales. Calcule además $F_{X,Y}(x, y)$.

296. Sean a y b dos constantes positivas. Calcule las densidades marginales del vector (X, Y) con función de densidad uniforme en la región que aparece en la Figura 3.12.

Distribución condicional

297. Demuestre que la función de distribución condicional $x \mapsto F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(u|y) du$ es efectivamente una función de distribución univariada.

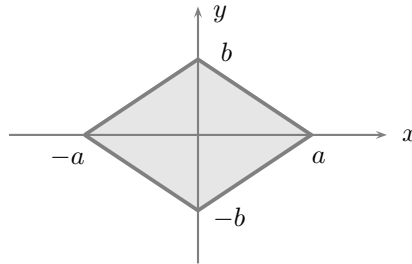


Figura 3.12:

298. Demuestre que la función de densidad condicional $x \mapsto f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y)$ es efectivamente una función de densidad univariada. En el caso absolutamente continuo compruebe además que $f_{X|Y}(x|y) = \partial/\partial x F_{X|Y}(x|y)$.
299. LA DISTRIBUCIÓN EXPONENCIAL NO TIENE MEMORIA. Sea X con distribución $\exp(\lambda)$ y sea $t > 0$ fijo. Demuestre que la distribución condicional de $X - t$, dado que $X \geq t$, sigue siendo $\exp(\lambda)$.
300. Calcule las funciones condicionales $f_{X|Y}(x|y)$ y $F_{X|Y}(x|y)$, para las siguientes funciones de densidad.
- $f(x,y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
 - $f(x,y) = 4xy$, para $0 < x, y < 1$.
 - $f(x,y) = 24x(1-x-y)$, para $x, y > 0$ y $x+y < 1$.
 - $f(x,y) = (x+2y)/4$, para $0 < x < 2$ y $0 < y < 1$.
 - $f(x,y) = 2(4x+y)/5$, para $0 < x, y < 1$.
 - $f(x,y) = 1/x$, para $0 < y < x < 1$.
 - $f(x,y) = xy/25$, para $0 < x < 2$, $0 < y < 5$.
301. Calcule las funciones condicionales $F_{X|Y}(x|y)$ y $f_{X|Y}(x|y)$, para las siguientes funciones de distribución conjunta.
- $F(x,y) = (1 - e^{-x})(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y)$, para $x \geq 0$.

b) $F(x, y) = 1 - e^{-x} - e^{-y} + e^{-x-y}$, para $x, y \geq 0$.

302. Se hacen tres lanzamientos de una moneda equilibrada cuyos resultados llamaremos cara y cruz. Sea X la variable que denota el número de caras que se obtienen en los dos primeros lanzamientos, y sea Y la variable que denota el número de cruces en los dos últimos lanzamientos. Calcule $f_{X,Y}(x, y)$, $f_X(x)$, $f_Y(y)$ y $f_{Y|X}(y|x)$ para $x = 0, 1, 2$.
303. Sea (X, Y) un vector con función de densidad $f(x, y) = (x + y)/8$, para $0 \leq x, y \leq 2$, con gráfica como se muestra en la Figura 3.13. Compruebe que $f(x, y)$ es una función de densidad y calcule

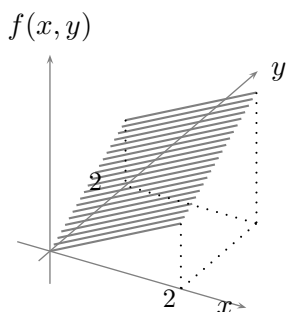


Figura 3.13: Función de densidad $f(x, y) = (x + y)/8$, para $x, y \in [0, 2]$.

- | | |
|----------------------|-------------------------|
| a) $f_X(x)$. | h) $F_{X Y}(x y)$. |
| b) $f_Y(y)$. | i) $F_{Y X}(y x)$. |
| c) $F_{X,Y}(x, y)$. | j) $P(Y > X)$. |
| d) $F_X(x)$. | k) $P(X > 1 Y < 1)$. |
| e) $F_Y(y)$. | l) $P(X > 1)$. |
| f) $f_{X Y}(x y)$. | m) $P(X + Y > 1)$. |
| g) $f_{Y X}(y x)$. | n) $P(X - Y > 1)$. |

304. Sea (X, Y) un vector con función de densidad $f(x, y) = 8xy$, para $0 < x < y < 1$. Grafique y compruebe que $f(x, y)$ es una función de densidad. Calcule además

- | | |
|----------------------|-----------------------------|
| a) $f_X(x)$. | h) $F_{X Y}(x y)$. |
| b) $f_Y(y)$. | i) $F_{Y X}(y x)$. |
| c) $F_{X,Y}(x, y)$. | j) $P(Y < 1/2, X < 1/2)$. |
| d) $F_X(x)$. | k) $P(Y > 1/2 X > 1/2)$. |
| e) $F_Y(y)$. | l) $P(XY < 1)$. |
| f) $f_{X Y}(x y)$. | m) $P(X + Y < 1)$. |
| g) $f_{Y X}(y x)$. | n) $P(X - Y < 1)$. |

305. Sea (X, Y) un vector con función de densidad $f(x, y) = \frac{1}{2}(x+y)e^{-x-y}$, para $x, y > 0$, cuya gráfica aparece en la Figura 3.14. Compruebe que $f(x, y)$ es una función de densidad y calcule

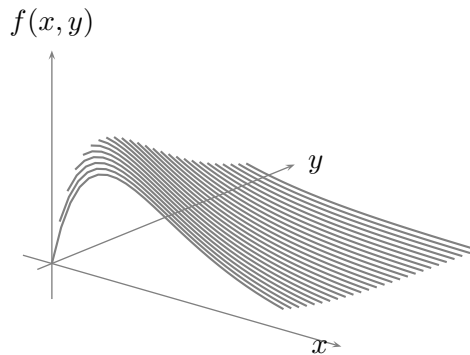


Figura 3.14: Función de densidad $f(x, y) = \frac{1}{2}(x+y)e^{-x-y}$, para $x, y > 0$.

- | | |
|---------------------|--------------------------------|
| a) $f_X(x)$. | h) $F_{X Y}(x y)$. |
| b) $f_Y(y)$. | i) $F_{Y X}(y x)$. |
| c) $F_{X,Y}(x,y)$. | j) $P(0 < X < 1, 0 < Y < 1)$. |
| d) $F_X(x)$. | k) $P(Y > 2 X < 1)$. |
| e) $F_Y(y)$. | l) $P(XY < 1)$. |
| f) $f_{X Y}(x y)$. | m) $P(X + Y > 1)$. |
| g) $f_{Y X}(y x)$. | n) $P(X - Y < 1)$. |

306. Sea (X, Y) un vector con función de densidad $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$, cuya gráfica se muestra en la Figura 3.15.

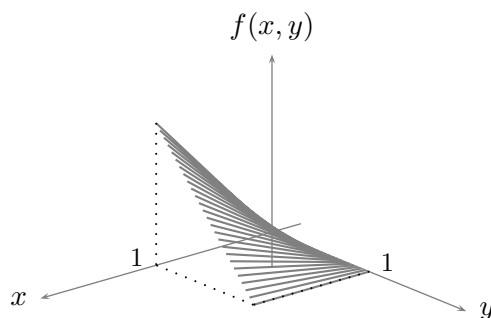


Figura 3.15: Función de densidad $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

Compruebe que $f(x, y)$ es efectivamente una función de densidad y calcule

- a) $f_X(x)$. h) $F_{X|Y}(x|y)$.
 b) $f_Y(y)$. i) $F_{Y|X}(y|x)$.
 c) $F_{X,Y}(x,y)$. j) $P(X > 1/2)$.
 d) $F_X(x)$. k) $P(1/4 < Y < 3/4 \mid X < 1/2)$.
 e) $F_Y(y)$. l) $P(Y > X^2)$.
 f) $f_{X|Y}(x|y)$. m) $P(2X - Y > 1)$.
 g) $f_{Y|X}(y|x)$. n) $P(|X - 2Y| < 1)$.

307. Sea (X, Y) un vector con función de densidad $f(x, y) = 3y$, para $0 < x < y < 1$. Compruebe que $f(x, y)$ es efectivamente una función de densidad y calcule

- a) $P(X + Y < 1/2)$.
 b) $f_X(x)$ y $f_Y(y)$.
 c) $E(Y)$ y $E(Y \mid X = x)$.

308. Sea (X, Y) un vector con distribución uniforme en el conjunto $\{1, \dots, 6\} \times \{1, \dots, 6\}$. Calcule

- a) $P(X = Y)$.
 b) $P(X + Y \leq 6)$.
 c) $f_X(x)$ y $f_Y(y)$.
 d) $E(X \mid X + Y = 6)$.

309. Sea (X, Y) un vector con función de densidad dada por la siguiente tabla

$x \backslash y$	-1	0	1
1	.3	.05	.05
2	.05	.2	.05
3	.1	.1	.1

Calcule

- a) $P(X = 2)$, $P(X + Y = 1)$ y $P(Y \leq X)$.

- b) $f_X(x)$ y $f_Y(y)$.
 - c) $f_{Y|X}(y|x)$ para $x = 1, 2, 3$.
 - d) $E(Y|X = x)$ para $x = 1, 2, 3$.
310. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Demuestre que la distribución condicional de X dado que $X + Y = u$, es uniforme en el intervalo $(0, u)$.
311. Sean A y B dos eventos con probabilidad positiva y sea X una variable con esperanza finita. Demuestre o proporcione un contraejemplo.
- a) Si $A \subseteq B$, entonces $E(X|A) \leq E(X|B)$.
 - b) $E(X|A) \leq E(X)$.

Independencia de variables aleatorias

312. Sean X y Y variables aleatorias discretas con valores en los conjuntos $\{x_1, x_2, \dots\}$ y $\{y_1, y_2, \dots\}$, respectivamente. Demuestre que X y Y son independientes si, y sólo si, para cualesquiera valores de los índices $i, j = 1, 2, \dots$ $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$.
313. Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad $f_{X,Y}(x, y)$. Demuestre que las variables X y Y son independientes si, y sólo si, para casi todo par de números x y y se cumple $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.
314. Demuestre la variable aleatoria constante es independiente de cualquier otra variable aleatoria. Inversamente, suponga que X es independiente de cualquier otra variable aleatoria, demuestre que X es constante.
315. Demuestre que los eventos A y B son independientes si, y sólo si, las variables aleatorias indicadoras 1_A y 1_B lo son.

316. Demuestre que si tres variables aleatorias son independientes, entonces cualesquiera dos de ellas lo son. Más generalmente, demuestre que cualquier subconjunto finito de un conjunto de variables aleatorias independientes también lo es.

317. Sean X y Y independientes. Demuestre que cada uno de los siguientes pares de variables aleatorias también son independientes. El siguiente ejercicio generaliza este resultado.

a) X y $-Y$. b) $|X|$ y $|Y|$.

318. Sean X_1, \dots, X_n independientes, y sean $g_1, \dots, g_n : \mathbb{R} \rightarrow \mathbb{R}$ funciones Borel medibles. Demuestre que las variables $g_1(X_1), \dots, g_n(X_n)$ son independientes.

319. Demuestre que las variables aleatorias X_1, \dots, X_n son independientes si, y sólo si, para cualquier vector (x_1, \dots, x_n) en \mathbb{R}^n se cumple

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

320. Sean X_1, \dots, X_n independientes, y sea $1 \leq k < n$. Sean $g : \mathbb{R}^k \rightarrow \mathbb{R}$ y $h : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$ funciones Borel medibles. Demuestre que las variables aleatorias $g(X_1, \dots, X_k)$ y $h(X_{k+1}, \dots, X_n)$ son independientes.

321. Sean X y Y dos variables aleatorias independientes. Recuerde las definiciones $X^+ = \max\{0, X\}$ y $X^- = -\min\{0, X\}$. Demuestre que cada uno de los siguientes pares de variables aleatorias también son independientes.

a) X^+ y Y^+ . b) X^+ y Y^- . c) X^- y Y^+ . d) X^- y Y^- .

322. Determine si las siguientes son funciones de densidad de variables aleatorias independientes.

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = 2x$, para $0 < x, y < 1$.

c) $f(x, y) = 2e^{-x-y}$, para $0 < x < y$.

- d) $f(x, y) = e^{-x-y}$, para $x, y > 0$.
 e) $f(x, y) = 3(x^2 + y^2)/8$, para $x, y \in [-1, 1]$.
 f) $f(x, y) = 3x(1 - xy)$, para $0 < x, y < 1$.

323. Determine si las siguientes son funciones de distribución de variables aleatorias independientes.

- a) $F(x, y) = (1 - e^{-x})(1 - e^{-y})$, para $x, y > 0$.
 b) $F(x, y) = (1 - e^{-x^2})(1 - e^{-y^2})$, para $x, y > 0$.

324. Demuestre que X y Y son independientes si, y sólo si, cualquiera de las siguientes condiciones se cumple: Para cada par de números reales x y y ,

- a) $P(X > x, Y > y) = P(X > x)P(Y > y)$.
 b) $P(X \leq x, Y > y) = P(X \leq x)P(Y > y)$.
 c) $P(X > x, Y \leq y) = P(X > x)P(Y \leq y)$.

325. Demuestre que X y Y son independientes si, y sólo si, para cualesquiera números reales $a < b$ y $c < d$,

$$P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d).$$

326. Diga falso o verdadero. Demuestre en cada caso.

- a) X, Y independientes $\Leftrightarrow X, Y^2$ independientes.
 b) X, Y independientes $\Leftrightarrow X^2, Y^2$ independientes.
 c) X, Y independientes $\Leftrightarrow X + Y, Y$ independientes.
 d) X, Y independientes $\Leftrightarrow X + Y, X - Y$ independientes.
 e) X, Y independientes $\Leftrightarrow XY, Y$ independientes.
 f) X, Y, Z independientes $\Leftrightarrow X + Y, Z$ independientes.
 g) X, Y, Z independientes $\Leftrightarrow XY, Z$ independientes.

327. Sean X y Y independientes ambas con distribución normal estándar. Demuestre que $Z = aX + bY + c$ tiene distribución normal cuando $ab \neq 0$. Encuentre la esperanza y varianza de Z .
328. Sean X_1, \dots, X_n variables aleatorias independientes cada una con distribución $\text{Ber}(p)$. Calcule $P(X_1 + \dots + X_n = k)$ para $k = 0, 1, \dots, n$.
329. Sean X y Y independientes ambas con distribución $\text{unif}\{1, \dots, n\}$. Encuentre la distribución del vector $(U, V) = (X + Y, X - Y)$. Determine además si las variables U y V son independientes.
330. Sean X y Y independientes con valores enteros naturales y con esperanza finita. Demuestre que

$$E(\text{mín}\{X, Y\}) = \sum_{n=1}^{\infty} P(X \geq n)P(Y \geq n).$$

331. Sean X y Y independientes con distribución Poisson de parámetros λ_1 y λ_2 respectivamente. Demuestre que la distribución condicional de X dado que $X + Y = n$ es $\text{bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.
332. Encuentre la función de densidad de $X + Y$ cuando X y Y son independientes con distribución uniforme en los conjuntos $\{0, 1, \dots, n\}$ y $\{0, 1, \dots, m\}$ respectivamente.
333. Sean X_1, \dots, X_n independientes con distribución $\text{geo}(p)$. Demuestre que la variable $X_1 + \dots + X_n$ tiene distribución $\text{bin neg}(n, p)$.
334. Sean X y Y independientes. Encuentre la función de distribución de Z en términos de $F_X(x)$ y $F_Y(y)$ cuando
 a) $Z = \text{máx}\{X, Y\}$. b) $Z = \text{mín}\{X, Y\}$.
335. Sean X y Y independientes ambas con distribución $\text{exp}(\lambda)$, y sea a una constante. Calcule $P(\text{máx}\{X, Y\} \leq aX)$ y $P(\text{mín}\{X, Y\} \leq aX)$.
336. Sean X y Y independientes con distribución $\text{exp}(\lambda_1)$ y $\text{exp}(\lambda_2)$ respectivamente. Demuestre que $P(Y > X) = \lambda_1/(\lambda_1 + \lambda_2)$.

337. Sean X y Y independientes e idénticamente distribuidas. Demuestre que $P(Y > X) = 1/2$.
338. Sean X y Y variables independientes con distribución exponencial con parámetros λ_1 y λ_2 respectivamente. Demuestre que $\min\{X, Y\}$ tiene distribución exponencial con parámetro $\lambda_1 + \lambda_2$, y que $P(X_1 = \min\{X_1, X_2\}) = \lambda_1/(\lambda_1 + \lambda_2)$. Este resultado puede extenderse al caso de n variables independientes exponenciales.
339. Usando la siguiente tabla, construya una función de densidad $f(x, y)$ de un vector discreto (X, Y) , distinta de la densidad uniforme, con la condición de que X y Y sean independientes.

$x \backslash y$	0	1
0	·	·
1	·	·

340. Sea (X, Y) un vector discreto con distribución de probabilidad uniforme en el conjunto $\{1, \dots, n\} \times \{1, \dots, m\}$, con n y m enteros positivos. Demuestre que X y Y son independientes.
341. Sea (X, Y) un vector con función de densidad $f(x, y) = c(1 - x)$, para $0 < x < y < 1$.
- Encuentre el valor de c que hace a $f(x, y)$ una función de densidad y grafique esta función.
 - Calcule $P(X + Y > 1)$ y $P(X \leq 1/2)$.
 - Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
 - Determine si X y Y son independientes.
342. Sea (X, Y) un vector aleatorio con función de densidad $f(x, y) = c/2^{x+y}$, para $x = 0, 1, 2$, y $y = 1, 2$. Encuentre el valor de la constante c y determine si X y Y son independientes. Calcule además las probabilidades $P(X = 1)$, $P(X = 2 | Y = 2)$ y $P(XY = 2)$.
343. Sea (X, Y) un vector aleatorio con función de densidad $f(x, y) = 2$, para $0 < x < y < 1$.

- a) Grafique y demuestre que $f(x, y)$ es una función de densidad.
 - b) Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
 - c) Determine si X y Y son independientes.
 - d) Calcule $P(Y > X)$ y $P(Y > X^2)$.
344. Sea (X, Y) un vector con función de densidad $f(x, y) = c|x + y|$, para $-1 < x, y < 1$.
- a) Encuentre el valor de la constante c que hace a $f(x, y)$ una función de densidad y grafique esta función.
 - b) Calcule $P(X > 0)$, $P(XY > 0)$ y $P(0 < X + Y < 1)$.
 - c) Encuentre las funciones de densidad marginales $f_X(x)$ y $f_Y(y)$.
 - d) Determine si X y Y son independientes.
345. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$, respectivamente. Demuestre que $X + Y$ tiene distribución $\text{bin}(n + m, p)$.
346. Sean X y Y independientes con distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Demuestre que $X + Y$ tiene distribución $\text{Poisson}(\lambda_1 + \lambda_2)$.
347. Sea (X, Y, Z) un vector aleatorio con función de densidad $f(x, y, z) = 8xyz$, para $0 < x, y, z < 1$.
- a) Compruebe que $f(x, y, z)$ es una función de densidad.
 - b) Calcule $P(X < Y < Z)$ y $P(X + Y + Z < 1)$.
 - c) Encuentre $f_{X,Y}(x, y)$, $f_{X,Z}(x, z)$ y $f_{Y,Z}(y, z)$.
 - d) Determine si X , Y y Z son independientes.
348. Sea (X, Y, Z) un vector aleatorio con función de densidad $f(x, y, z) = 24x$, para $0 < x < y < z < 1$.
- a) Compruebe que $f(x, y, z)$ es una función de densidad.
 - b) Calcule $P(X + Y < 1)$ y $P(Z - X > 1/2)$.

- c) Encuentre $f_{X,Y}(x, y)$, $f_{X,Z}(x, z)$ y $f_{Y,Z}(y, z)$.
 d) Determine si X , Y y Z son independientes.

349. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes cada una con distribución $\text{unif}(0, 1)$. Demuestre que para cualquier $\lambda > 0$,

$$\lim_{n \rightarrow \infty} P(\text{máx}\{X_1, \dots, X_n\} \leq 1 - \lambda/n) = e^{-\lambda}.$$

350. Sean X y Y independientes con distribución Poisson de parámetros λ_1 y λ_2 respectivamente. Demuestre que

$$E(X | X + Y = n) = n \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

351. Encuentre una distribución conjunta de dos variables aleatorias X y Y que no sean independientes y que Y tenga distribución marginal $\text{Ber}(p)$.

Esperanza de una función de un vector aleatorio

352. Demuestre que la condición $E(XY) = E(X)E(Y)$ no implica necesariamente que X y Y son independientes. Para ello considere cualquiera de los siguientes ejemplos.

$$a) f(x, y) = \begin{cases} 1/8 & \text{si } (x, y) = (1, 1), (1, -1), (-1, 1), (-1, -1), \\ 1/2 & \text{si } (x, y) = (0, 0), \\ 0 & \text{otro caso.} \end{cases}$$

$$b) f(x, y) = 3(x^2 + y^2)/8, \text{ para } x, y \in [-1, 1].$$

$$c) X \text{ con distribución uniforme en } \{-1, 0, 1\} \text{ y } Y = 1_{(X \neq 0)}.$$

353. Demuestre que si las variables X_1, \dots, X_n son independientes e integrables, entonces $E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n)$.
354. Sean X y Y independientes. Diga falso o verdadero justificando en cada caso.

a) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

b) $\text{Var}(X - Y) = \text{Var}(X) - \text{Var}(Y)$.

c) $\text{Var}(XY) = \text{Var}(X)\text{Var}(Y)$.

355. Sean X y Y variables aleatorias independientes con varianza finita. Demuestre que

$$\text{Var}(XY) = \text{Var}(X)\text{Var}(Y) + E^2(X)\text{Var}(Y) + E^2(Y)\text{Var}(X).$$

356. Sean X_1, \dots, X_n independientes con idéntica distribución y con esperanza finita. Demuestre que si x es tal que $f_{X_1+\dots+X_n}(x) \neq 0$, entonces

$$E(X_1 | X_1 + \dots + X_n = x) = \frac{x}{n}.$$

357. Sea (X, Y) un vector aleatorio discreto con función de densidad $f(x, y)$ dada por la siguiente tabla.

$x \backslash y$	-1	0	1
1	.1	.05	.1
2	.06	.2	.04
3	.1	.05	.3

a) Grafique $f(x, y)$ y compruebe que efectivamente se trata de una función de densidad conjunta.

b) Calcule y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas funciones son efectivamente de densidad.

c) Demuestre que X y Y no son independientes.

d) Calcule $E(XY)$ y $f_{X+Y}(u)$.

358. Sea (X, Y) un vector discreto con función de densidad dada por la siguiente tabla.

$x \backslash y$	2	4	6
1	2/18	3/18	1/18
2	3/18	5/18	1/18
3	1/18	1/18	1/18

- a) Grafique $f(x, y)$ y compruebe que efectivamente es una función de densidad conjunta.
- b) Calcule y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas son efectivamente funciones de densidad.
- c) Demuestre que X y Y no son independientes.
- d) Calcule $E(XY)$ y $f_{X+Y}(u)$.

359. Sea (X, Y) un vector aleatorio con función de densidad dada por

$$f(x, y) = \begin{cases} 8xy & \text{si } 0 < y < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

- a) Grafique $f(x, y)$ y compruebe que efectivamente es una función de densidad conjunta.
- b) Encuentre y grafique las densidades marginales $f_X(x)$ y $f_Y(y)$. Verifique que ambas son efectivamente funciones de densidad.
- c) Demuestre que X y Y no son independientes.
- d) Calcule $E(XY)$ y $f_{X+Y}(u)$.

Esperanza y varianza de un vector

360. Calcule la esperanza y varianza del vector aleatorio (X, Y) cuya función de densidad conjunta es

- a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.
- b) $f(x, y) = 4xy$, para $x, y \in [0, 1]$.

Covarianza

361. Sea a cualquier número real fijo. Encuentre variables aleatorias X y Y tales que $\text{Cov}(X, Y) = a$,

362. Diga falso o verdadero. Demuestre en cada caso.

- a) $X \geq 0, Y \geq 0 \Rightarrow \text{Cov}(X, Y) \geq 0$.
- b) $\text{Cov}(X, Y) = 0, \text{Cov}(Y, Z) = 0 \Rightarrow \text{Cov}(X, Z) = 0$.
- c) $\text{Cov}(X, Y) > 0, \text{Cov}(Y, Z) > 0 \Rightarrow \text{Cov}(X, Z) > 0$.
- d) $\text{Cov}(X, Y) = a, \text{Cov}(Y, Z) = a \Rightarrow \text{Cov}(X, Z) = a$.

363. Diga falso o verdadero. Demuestre en cada caso.

- a) $\text{Cov}(X, Y) \geq 0$.
- b) $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$, con a, b constantes.
- c) $\text{Cov}(X, aY + b) = a \text{Cov}(X, Y) + b$, con a, b constantes.

364. Demuestre que

- a) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.
- b) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- c) $\text{Cov}(X, X) = \text{Var}(X)$.
- d) $\text{Cov}(X, -X) = -\text{Var}(X)$.
- e) $\text{Cov}(aX + b, Y) = a \text{Cov}(X, Y)$, con a, b constantes.
- f) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

365. Demuestre que la condición $\text{Cov}(X, Y) = 0$ no es suficiente para concluir que X y Y son independientes. En el texto se proporciona un ejemplo para un vector discreto, construya ahora un ejemplo para un vector continuo.

366. Demuestre que $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \text{Cov}(X, Y)$.

367. Demuestre que

$$a) \text{Var}(X_1 + \cdots + X_n) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{j < k} \text{Cov}(X_j, X_k).$$

$$b) \operatorname{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \operatorname{Cov}(X_i, Y_j).$$

368. Sean X_1, \dots, X_n independientes y con varianza finita. Demuestre que

$$\operatorname{Var}(X_1 + \dots + X_n) = \sum_{k=1}^n \operatorname{Var}(X_k).$$

369. Sean X_1, \dots, X_n independientes y con idéntica distribución. Defina $\bar{X} = (X_1 + \dots + X_n)/n$. Demuestre que para cada $k = 1, \dots, n$, $\operatorname{Cov}(X_k - \bar{X}, \bar{X}) = 0$.

370. Sea (X, Y) con distribución uniforme en el conjunto $\{1, \dots, n\} \times \{1, \dots, n\}$. Demuestre que $\operatorname{Cov}(X, Y) = 0$.

371. Sea (X, Y) con distribución uniforme en el conjunto $(a, b) \times (c, d)$. Demuestre que $\operatorname{Cov}(X, Y) = 0$.

372. Calcule la covarianza de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

$x \backslash y$	-1	0	1
-1	1/12	2/12	3/12
1	3/12	2/12	1/12

373. Calcule la covarianza de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

$x \backslash y$	1	2	3
2	.2	.05	.15
4	.05	.1	.15
6	.05	.1	.15

374. Calcule la covarianza de X y Y , cuya función de densidad conjunta es

$$a) f(x, y) = \frac{1}{ab}, \text{ para } 0 < x < a, 0 < y < b.$$

b) $f(x, y) = 3x^2y$, para $-1 < x < 1$, $0 < y < 1$.

c) $f(x, y) = e^{-x}/2$, para $|y| < x$.

d) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

375. Sea (X, Y) un vector con distribución normal $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $\text{Cov}(X, Y) = \rho \sigma_X \sigma_Y$.

Coefficiente de correlación

376. Demuestre nuevamente que $-1 \leq \rho(X, Y) \leq 1$, ahora a partir de la desigualdad de Cauchy-Schwarz.

377. Diga falso o verdadero. Demuestre en cada caso.

a) $\rho(X, Y) = 0$, $\rho(Y, Z) = 0 \Rightarrow \rho(X, Z) = 0$.

b) $\rho(X, Y) > 0$, $\rho(Y, Z) > 0 \Rightarrow \rho(X, Z) > 0$.

c) $\rho(X, Y) < 0$, $\rho(Y, Z) < 0 \Rightarrow \rho(X, Z) < 0$.

d) $\rho(X, Y) = 1$, $\rho(Y, Z) = 1 \Rightarrow \rho(X, Z) = 1$.

e) $\rho(X, Y) = -1$, $\rho(Y, Z) = -1 \Rightarrow \rho(X, Z) = -1$.

f) $\rho(X, Y)\rho(Y, Z) = -1 \Rightarrow \rho(X, Z) = -1$.

g) $\rho(X, Y) = a$, $\rho(Y, Z) = a \Rightarrow \rho(X, Z) = a$.

378. Diga falso verdadero. Demuestre en cada caso.

a) $\rho(X, Y) = \rho(Y, X)$.

b) $\rho(X + a, Y) = \rho(X, Y)$, a constante.

c) $\rho(aX + b, Y) = a\rho(X, Y) + b$, a, b constantes.

379. Sea a un número cualquiera en $[-1, 1]$. Encuentre variables aleatorias X y Y tales que $\rho(X, Y) = a$.

380. Sean X y Y independientes con distribución $\text{Ber}(p)$ con $p = 1/2$. Demuestre que el coeficiente de correlación entre $X + Y$ y $|X - Y|$ es cero, y sin embargo estas variables aleatorias no son independientes.

381. Sea X con distribución normal estándar. Demuestre que el coeficiente de correlación entre X y X^2 es cero, y sin embargo estas variables no son independientes. Este resultado puede extenderse al caso en el que la distribución de X cumple la condición $E(X) = E(X^3) = 0$.
382. Sea X una variable aleatoria y sean a y b constantes. Demuestre que
- $\rho(X, X) = 1$.
 - $\rho(X, -X) = -1$.
 - $\rho(X, aX + b) = \text{signo}(a)$.
383. Demuestre que $\rho(aX + b, cY + d) = \text{signo}(ac) \cdot \rho(X, Y)$, en donde $ac \neq 0$. Recuerde que

$$\text{signo}(x) = \begin{cases} +1 & \text{si } x > 0, \\ -1 & \text{si } x < 0, \\ 0 & \text{si } x = 0. \end{cases}$$

384. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

$x \backslash y$	1	2
0	1/8	1/4
1	1/2	1/8

385. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta está dada por la siguiente tabla.

$x \backslash y$	1	2	3
2	1/9	1/9	1/9
4	1/9	1/9	1/9
6	1/9	1/9	1/9

386. Calcule el coeficiente de correlación de X y Y con distribución conjunta uniforme en el conjunto

a) $\{1, \dots, n\} \times \{1, \dots, n\}$.

b) $[-1, 1] \times [-1, 1]$.

387. Sea X con distribución $\text{bin}(n, p)$ y sea $Y = n - X$. Demuestre que $\text{Cov}(X, Y) = -np(1 - p)$, y por lo tanto $\rho(X, Y) = -1$.

388. Calcule el coeficiente de correlación de X y Y cuya función de densidad conjunta es

a) $f(x, y) = \frac{1}{2} \text{sen}(x + y)$, para $x, y \in [0, \pi/2]$.

b) $f(x, y) = e^{-x}/2$, para $|y| < x$.

c) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

389. Sea (X, Y) un vector con distribución normal $N(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $\rho(X, Y) = \rho$.

Distribución multinomial

390. Demuestre que la función de densidad de la distribución multinomial efectivamente lo es.

391. Sea (X_1, \dots, X_{k-1}) un vector con distribución multinomial de parámetros (n, p_1, \dots, p_{k-1}) . Demuestre que cada coordenada X_i tiene distribución marginal $\text{bin}(n, p_i)$, para $i = 1, \dots, k - 1$.

392. Sea $X = (X_1, \dots, X_{k-1})$ un vector con distribución multinomial de parámetros (n, p_1, \dots, p_{k-1}) . Demuestre que $E(X) = (np_1, \dots, np_{k-1})$ y que

$$[\text{Var}(X)]_{ij} = \begin{cases} np_i(1 - p_i) & \text{si } i = j, \\ -np_i p_j & \text{si } i \neq j. \end{cases}$$

Observe que en el caso $i \neq j$, el signo negativo en la covarianza indica que cuando una variable crece la otra decrece.

Distribución hipergeométrica multivariada

393. Demuestre que la función de densidad de la distribución hipergeométrica multivariada efectivamente lo es.
394. Sea (X_1, \dots, X_k) un vector con distribución hipergeométrica multivariada con parámetros (N, N_1, \dots, N_k, n) . Demuestre que cada coordenada X_i tiene distribución hipergeométrica univariada con parámetros (N, N_i, n) , para $i = 1, \dots, k$.
395. Sea $X = (X_1, \dots, X_k)$ con distribución hipergeométrica multivariada con parámetros (N, N_1, \dots, N_k, n) . Demuestre que $E(X) = (nN_1/N, \dots, nN_k/N)$, y que

$$[\text{Var}(X)]_{ij} = \begin{cases} n \cdot \frac{N_i}{N} \cdot \frac{N - N_i}{N} \cdot \frac{N - n}{N - 1} & \text{si } i = j, \\ n \cdot \frac{N_i}{N} \cdot \frac{N_j}{N} \cdot \frac{n - N}{N - 1} & \text{si } i \neq j. \end{cases}$$

Distribución normal bivariada

396. Demuestre que la función de densidad de la distribución normal bivariada efectivamente lo es.
397. Sea (X, Y) un vector con distribución normal $N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Demuestre que X tiene distribución marginal $N(\mu_1, \sigma_1^2)$, y Y tiene distribución marginal $N(\mu_2, \sigma_2^2)$. Véase el siguiente ejercicio para verificar que el recíproco de este resultado es falso.
398. Sea $f(x, y)$ la función de densidad normal bivariada estándar con $\rho = 0$. Defina

$$g(x, y) = \begin{cases} 2f(x, y) & \text{si } xy < 0, \\ 0 & \text{si } xy \geq 0. \end{cases}$$

Demuestre que $g(x, y)$ es una función de densidad bivariada que no es normal pero cuyas densidades marginales son normales estándar.

399. Sea (X, Y) un vector con distribución normal $(\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2, \rho)$. Demuestre que $E(X) = (\mu_X, \mu_Y)$, y

$$\text{Var}(X, Y) = \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

400. Sea (X, Y) un vector con distribución normal $N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. Demuestre que la distribución condicional de Y dado que $X = x$ es normal con media $\mu_2 + \rho(x - \mu_1)\sigma_2/\sigma_1$ y varianza $\sigma_2^2(1 - \rho^2)$, y que la distribución condicional de X dado que $Y = y$ es normal con media $\mu_1 + \rho(y - \mu_2)\sigma_1/\sigma_2$ y varianza $\sigma_1^2(1 - \rho^2)$.
401. Demuestre que a través del cambio de variable $(u, v) = ((x - \mu_X)/\sigma_X, (y - \mu_Y)/\sigma_Y)$, la función de densidad normal bivariada se puede escribir como sigue

$$f(u, v) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)}(u - \rho v)^2 - \frac{1}{2}v^2\right).$$

Sea $c > 0$ y defina $k = 1/(2\pi c\sqrt{1 - \rho^2})$. Demuestre ahora que las líneas de contorno $f(u, v) = c$ son las elipses

$$\frac{(u - \rho v)^2}{\ln k^{2(1 - \rho^2)}} + \frac{v^2}{\ln k^2} = 1.$$

Cuando $\rho = 0$ las elipses se reducen al círculo $u^2 + v^2 = \ln k^2$.

CAPÍTULO 4

Esperanza condicional

En este capítulo se presenta una breve introducción al concepto de esperanza condicional de una variable aleatoria respecto de una σ -álgebra, y se estudian algunas de sus propiedades elementales. Consideraremos que se cuenta con un espacio de probabilidad base (Ω, \mathcal{F}, P) , y que \mathcal{G} es una sub σ -álgebra de \mathcal{F} . Hemos definido antes la esperanza de una variable aleatoria X como la integral de Riemann-Stieltjes

$$E(X) = \int_{-\infty}^{\infty} x dF_X(x),$$

sin embargo, para hacer la notación más simple en este capítulo, es conveniente en algunas ocasiones adoptar la notación de la teoría de la medida y denotar la esperanza de una variable aleatoria X mediante la siguiente integral

$$E(X) = \int_{\Omega} X dP.$$

Esto corresponde a la integral de Lebesgue de la función medible X respecto de la medida de probabilidad P .

Recordemos que si se conoce la distribución de un vector (X, Y) y se toma un valor y tal que $f_Y(y) \neq 0$, la esperanza condicional de X dado $Y = y$ es

la función

$$y \mapsto E(X | Y = y) = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y),$$

cuando $f_Y(y) \neq 0$. De manera análoga, si A es un evento con probabilidad positiva y X es una variable aleatoria integrable, la esperanza condicional de X dado A es el número

$$E(X | A) = \int_{-\infty}^{\infty} x dF_{X|A}(x),$$

en donde $F_{X|A}(x) = P(X \leq x | A) = P(X \leq x, A)/P(A)$. La esperanza condicional que definiremos en este capítulo generaliza estos conceptos.

4.1. Esperanza condicional

He aquí la definición general. Es importante hacer énfasis que la esperanza condicional, a pesar de su nombre, no es un número, aunque puede serlo, sino una variable aleatoria.

DEFINICIÓN. (ESPERANZA CONDICIONAL). Sea X una variable aleatoria con esperanza finita, y sea \mathcal{G} una sub- σ -álgebra de \mathcal{F} . La esperanza condicional de X dado \mathcal{G} , es una variable aleatoria denotada por $E(X | \mathcal{G})$, que cumple las siguientes tres propiedades.

- a) Es \mathcal{G} -medible.
- b) Tiene esperanza finita.
- c) Para cualquier evento G en \mathcal{G} ,

$$\int_G E(X | \mathcal{G}) dP = \int_G X dP. \quad (4.1)$$

Parte de la dificultad para entender esta definición general es que no se proporciona una fórmula explícita para esta variable aleatoria sino únicamente las propiedades que cumple. El objetivo de este capítulo es encontrar el significado de esta variable aleatoria, interpretar su significado y explicar su relación con el concepto de esperanza condicional elemental, $E(X | Y = y)$, mencionado antes. Haremos lo anterior principalmente en el caso cuando la σ -álgebra \mathcal{G} es generada por una variable aleatoria discreta.

Usando el teorema de Radon-Nikodym (véase por ejemplo [5]), puede demostrarse que esta variable aleatoria existe y es única casi seguramente, esto significa que si existe otra variable aleatoria que cumple las tres propiedades de la definición anterior, entonces con probabilidad uno coincide con $E(X | \mathcal{G})$. En lo sucesivo cuando se establezca que esta variable aleatoria es igual a alguna otra variable, la igualdad debe entenderse en el sentido casi seguro, es decir, que la igualdad se verifica con probabilidad uno.

NOTACIÓN.

- a) Cuando la σ -álgebra \mathcal{G} es igual a $\sigma(Y)$, para alguna variable aleatoria Y , la esperanza condicional se escribe simplemente como $E(X | Y)$, en lugar de $E(X | \sigma(Y))$.
- b) Si A es un evento, entonces la esperanza condicional $E(1_A | \mathcal{G})$ se denota por $P(A | \mathcal{G})$.

En la siguiente sección estudiaremos con más detalle la variable $E(X | Y)$ cuando Y es una variable aleatoria discreta.

4.2. Esperanza condicional: caso discreto

Sean X y Y dos variables aleatorias. Suponga que X tiene esperanza finita y que Y es discreta con posibles valores y_1, y_2, \dots . La esperanza condicional de X dado el evento $(Y = y_j)$ es el número $E(X | Y = y_j)$. Este valor depende naturalmente del evento $(Y = y_j)$, y podemos considerar que es una función de los posibles valores y_j , o bien que tenemos una función definida sobre el espacio muestral de la siguiente forma: Si ω es tal que $Y(\omega) = y_j$, entonces

$$\omega \mapsto E(X | Y)(\omega) = E(X | Y = y_j).$$

Observe que $E(X | Y)$ toma a lo sumo tantos valores distintos como lo hace la variable Y . Globalmente se puede escribir esta función en términos de funciones indicadoras como sigue

$$E(X | Y)(\omega) = \sum_{j=1}^{\infty} E(X | Y = y_j) 1_{(Y=y_j)}(\omega).$$

De este modo se construye la función $E(X | Y) : \Omega \rightarrow \mathbb{R}$, que resulta ser una variable aleatoria, y corresponde a un caso particular de la definición general enunciada antes. Demostraremos esto a continuación.

PROPOSICIÓN. (ESPERANZA CONDICIONAL, CASO DISCRETO). Sea X una variable aleatoria integrable, y sea Y discreta con valores y_1, y_2, \dots . La función $E(X | Y) : \Omega \rightarrow \mathbb{R}$ dada por

$$\omega \mapsto E(X | Y)(\omega) = E(X | Y = y_j) \quad \text{si } Y(\omega) = y_j,$$

es una variable aleatoria que cumple las siguientes propiedades.

- a) Es $\sigma(Y)$ -medible.
- b) Tiene esperanza finita.
- c) Para cualquier evento G en $\sigma(Y)$,

$$\int_G E(X | Y) dP = \int_G X dP. \quad (4.2)$$

Demostración.

- a) A través de sus posibles valores la variable aleatoria Y secciona el espacio muestral Ω en eventos disjuntos, a saber, $(Y = y_1), (Y = y_2), \dots$. La mínima σ -álgebra respecto de la cual la función Y es variable aleatoria es $\sigma(Y) = \sigma\{(Y = y_1), (Y = y_2), \dots\} \subseteq \mathcal{F}$. Como $E(X | Y)$ es constante en cada elemento de la partición, resulta que esta función es $\sigma(Y)$ -medible, y en consecuencia es verdaderamente una variable aleatoria.
- b) Tomando el evento G como Ω en la tercera propiedad se obtiene que tanto X como $E(X | Y)$ tienen la misma esperanza.
- c) Como cada elemento de $\sigma(Y)$ es una unión ajena de elementos de la forma $(Y = y_j)$, por propiedades de la integral es suficiente demos-

trar (4.2) para estos eventos simples. Tenemos entonces que

$$\begin{aligned}
 \int_{(Y=y_j)} E(X|Y)(\omega) dP(\omega) &= E(X|Y=y_j) P(Y=y_j) \\
 &= \int_{\Omega} X(\omega) dP(\omega|Y=y_j) P(Y=y_j) \\
 &= \int_{\Omega} X(\omega) dP(\omega, Y=y_j) \\
 &= \int_{(Y=y_j)} X(\omega) dP(\omega).
 \end{aligned}$$

□

Observe la diferencia entre $E(X|Y=y_j)$ y $E(X|Y)$. El primer término es un posible valor numérico del segundo término que es una variable aleatoria, sin embargo a ambas expresiones se les llama esperanza condicional. Veremos a continuación un caso particular de esta variable aleatoria. Demostraremos que la esperanza condicional puede verse como una generalización del concepto básico de probabilidad condicional, y también puede considerarse como una generalización del concepto de esperanza.

PROPOSICIÓN. Sea X con esperanza finita, y sean A y B eventos tales que $0 < P(B) < 1$. Entonces

1. $E(X|\{\emptyset, \Omega\}) = E(X)$.
2. $E(1_A|\{\emptyset, \Omega\}) = P(A)$.
3. $E(1_A|\{\emptyset, B, B^c, \Omega\}) = P(A|B)1_B + P(A|B^c)1_{B^c}$.

Demostración.

1. Esta igualdad se sigue del hecho que la variable $E(X|\mathcal{G})$ es medible respecto de \mathcal{G} , y de que cualquier función medible respecto de la

σ -álgebra $\{\emptyset, \Omega\}$ es constante. La tercera condición en la definición general de esperanza condicional implica que esta constante debe ser $E(X)$.

2. Esta igualdad es evidentemente un caso particular de la primera.
3. Observe que toda función medible respecto de la σ -álgebra \mathcal{G} dada por $\{\emptyset, B, B^c, \Omega\}$ es constante tanto en B como en B^c . Además,

$$\int_B E(1_A | \mathcal{G}) dP = \int_B 1_A dP = P(A \cap B).$$

Como la variable aleatoria $E(1_A | \mathcal{G})$ es constante en B , el lado izquierdo es igual a $E(1_A | \mathcal{G})(\omega) P(B)$, para cualquier ω en B . De donde se obtiene $E(1_A | \mathcal{G})(\omega) = P(A | B)$, para cualquier ω en B . El análisis es análogo al considerar el evento B^c , y de esto se obtiene la fórmula enunciada.

□

Observe en particular que la tercera propiedad dice que si la σ -álgebra \mathcal{G} es generada por la partición elemental $\{B, B^c\}$, entonces la esperanza condicional de 1_A es una variable aleatoria que toma dos valores: $P(A | B)$ sobre B , y $P(A | B^c)$ sobre B^c . El siguiente ejercicio es una generalización de este resultado.

EJERCICIO. Sea B_1, \dots, B_n una partición de Ω tal que cada uno de estos elementos tiene probabilidad estrictamente positiva. Demuestre que para cualquier evento A ,

$$E(1_A | \sigma\{B_1, \dots, B_n\}) = \sum_{i=1}^n P(A | B_i) 1_{B_i}.$$

▪

EJERCICIO. Encuentre una distribución conjunta de dos variables aleatorias

X y Y de tal forma que $E(X | Y)$ tenga distribución $\text{Ber}(p)$. ▪

La variable $E(X | \mathcal{G})$ puede interpretarse como la esperanza condicional de X dada la *información* de la σ -álgebra \mathcal{G} . Ilustraremos esto en el siguiente ejemplo.

EJEMPLO. Considere el experimento aleatorio de lanzar un dado equilibrado e intentar adivinar el resultado que se obtiene. Suponga, por ejemplo, que se apuesta a que se obtiene el número "2". Defina los eventos $A = \{2\}$, $B = \{2, 4, 6\}$ y la colección $\mathcal{G} = \{\emptyset, \Omega, B, B^c\}$. Esta σ -álgebra puede *distinguir* los resultados "Cae número par", evento B , y "Cae número impar", evento B^c . Entonces

$$\begin{aligned} E(1_A | \mathcal{G}) &= P(A | B) 1_B(\omega) + P(A | B^c) 1_{B^c}(\omega) \\ &= \frac{1}{3} 1_B(\omega) + 0 1_{B^c}(\omega) \\ &= \begin{cases} \frac{1}{3} & \text{si } \omega \in B, \\ 0 & \text{si } \omega \in B^c. \end{cases} \end{aligned}$$

De esta forma $E(1_A | \mathcal{G})$ es una función que reporta las probabilidades de ganar apostando por el número "2" en cada una de las dos situaciones que la σ -álgebra \mathcal{G} distingue: resultado par o impar.

Si se toma, en cambio, $\mathcal{G} = 2^\Omega$ con $\Omega = \{1, 2, 3, 4, 5, 6\}$, es decir, la σ -álgebra "distingue" totalmente el resultado del lanzamiento del dado, entonces definiendo los eventos $B_i = \{i\}$, para $i = 1, \dots, 6$, tenemos

$$E(1_A | \mathcal{G}) = \sum_{i=1}^6 P(A | B_i) 1_{B_i} = P(A | A) 1_A = 1_A.$$

Es decir, no hay sorpresas, cuando se sabe con precisión el resultado del experimento, conocemos obviamente la probabilidad de ganar apostando por el número "2", ésta es cero o uno. ▪

4.3. Algunas propiedades

Veremos ahora algunas propiedades generales de la esperanza condicional, otras propiedades se encuentran en la sección de ejercicios. Recordemos que las identidades o desigualdades que se establezcan para la esperanza condicional son válidas en el sentido casi seguro, es decir, con probabilidad uno se cumplen las afirmaciones. En un apéndice al final del texto se encuentra una lista más completa de propiedades de esta variable aleatoria.

PROPOSICIÓN. Sean X y Y variables aleatorias con esperanza finita y sea c una constante. Entonces

1. Si $X \geq 0$, entonces $E(X | \mathcal{G}) \geq 0$.
2. $E(cX + Y | \mathcal{G}) = cE(X | \mathcal{G}) + E(Y | \mathcal{G})$.
3. Si $X \leq Y$, entonces $E(X | \mathcal{G}) \leq E(Y | \mathcal{G})$.
4. $E(E(X | \mathcal{G})) = E(X)$.
5. Si X es \mathcal{G} -medible, entonces $E(X | \mathcal{G}) = X$ c.s.
En particular, $E(c | \mathcal{G}) = c$.
6. Si $\mathcal{G}_1 \subseteq \mathcal{G}_2$, entonces

$$E(E(X | \mathcal{G}_1) | \mathcal{G}_2) = E(E(X | \mathcal{G}_2) | \mathcal{G}_1) = E(X | \mathcal{G}_1).$$

Demostración.

1. Por contradicción, suponga que existe G en \mathcal{G} con probabilidad estrictamente positiva tal que $E(X | \mathcal{G}) \cdot 1_G < 0$. Entonces tomando esperanzas se obtiene $E(X \cdot 1_G) < 0$. Por otro lado, como $X \geq 0$, $E(X \cdot 1_G) \geq 0$.

2. Esta igualdad es consecuencia de la linealidad de la esperanza no condicional, junto con (4.1) y la propiedad de unicidad.
3. Esto consecuencia de la primera propiedad y la linealidad aplicadas a la variable $Y - X \geq 0$.
4. Esta propiedad se obtiene tomando $G = \Omega$ en la igualdad (4.1).
5. Si X es \mathcal{G} -medible, entonces X mismo cumple con las tres propiedades de la definición de esperanza condicional, por la unicidad se obtiene la igualdad casi segura.
6. Para todo $G \in \mathcal{G}_1 \subseteq \mathcal{G}_2$,

$$\int_G E(E(X | \mathcal{G}_1) | \mathcal{G}_2) dP = \int_G E(X | \mathcal{G}_1) dP = \int_G X dP.$$

Análogamente,

$$\int_G E(E(X | \mathcal{G}_2) | \mathcal{G}_1) dP = \int_G E(X | \mathcal{G}_2) dP = \int_G X dP.$$

□

En particular observe que la segunda propiedad dice que la esperanza condicional es lineal, mientras que la cuarta propiedad establece que las variables aleatorias X y $E(X | \mathcal{G})$ tienen la misma esperanza, o en términos de información, la σ -álgebra trivial $\{\emptyset, \Omega\}$ realmente no proporciona ninguna información adicional del experimento aleatorio y por lo tanto la esperanza se calcula directamente sobre la variable aleatoria.

EJERCICIO. Demuestre las siguientes desigualdades:

$$a) |E(X | \mathcal{G})| \leq E(|X| | \mathcal{G}). \quad b) E|E(X | \mathcal{G})| \leq E(|X|). \quad \cdot$$

EJERCICIO. Sean X y Y independientes cada una con distribución $\text{Ber}(p)$. Encuentre $E(X | X + Y)$. ▪

Una introducción a la esperanza condicional ligeramente más completa a la

presentada en esta sección, aunque también sencilla y breve, puede encontrarse en [24]. Un tratamiento más completo y riguroso puede consultarse por ejemplo en [18] o [31].

4.4. Varianza condicional

Usando la esperanza condicional se puede obtener la varianza condicional de una variable aleatoria respecto de una σ -álgebra de la siguiente forma.

DEFINICIÓN. (VARIANZA CONDICIONAL). Sea X con segundo momento finito, y sea \mathcal{G} una sub- σ -álgebra de \mathcal{F} . La varianza condicional de X dado \mathcal{G} , denotada por $\text{Var}(X | \mathcal{G})$, se define como la variable aleatoria

$$\text{Var}(X | \mathcal{G}) = E[(X - E(X|\mathcal{G}))^2 | \mathcal{G}].$$

Nuevamente hacemos énfasis en que la varianza condicional no es necesariamente un número sino una variable aleatoria en general, \mathcal{G} -medible por definición, y como en el caso no condicional, es no negativa. Después de un cálculo sencillo se puede comprobar que se reconstruye la varianza no condicional en el siguiente caso particular: $\text{Var}(X | \{\emptyset, \Omega\}) = \text{Var}(X)$.

Demostraremos a continuación algunas propiedades elementales de esta variable aleatoria. Otras propiedades se encuentran en la sección de ejercicios.

PROPOSICIÓN. Sean X y Y con varianza finita, y sea c una constante. Entonces

1. $\text{Var}(X | \mathcal{G}) \geq 0$.
2. $\text{Var}(c | \mathcal{G}) = 0$.
3. $\text{Var}(cX | \mathcal{G}) = c^2 \text{Var}(X | \mathcal{G})$.
4. $\text{Var}(X + c | \mathcal{G}) = \text{Var}(X | \mathcal{G})$.
5. En general, $\text{Var}(X + Y | \mathcal{G}) \neq \text{Var}(X | \mathcal{G}) + \text{Var}(Y | \mathcal{G})$.
6. $\text{Var}(X | \mathcal{G}) = E(X^2 | \mathcal{G}) - E^2(X | \mathcal{G})$.
7. $\text{Var}(X) = E[\text{Var}(X | \mathcal{G})] + \text{Var}[E(X | \mathcal{G})]$.

Demostración.

- 1.–4. Estas propiedades son una consecuencia inmediata de las propiedades ya demostradas de la esperanza condicional.
5. Nuevamente es suficiente tomar $Y = X$ para verificar la no igualdad.
6. Esta igualdad se obtiene a partir de la definición al desarrollar el cuadrado y utilizar las propiedades de linealidad de la esperanza condicional.
7. Tomando esperanza en la igualdad previa se obtiene

$$E[\text{Var}(X | \mathcal{G})] = E(X^2) - E[E^2(X | \mathcal{G})].$$

Por otro lado,

$$\begin{aligned} \text{Var}[E(X | \mathcal{G})] &= E[E^2(X | \mathcal{G})] - E^2[E(X | \mathcal{G})] \\ &= E[E^2(X | \mathcal{G})] - E^2(X). \end{aligned}$$

Sumando estas últimas dos expresiones se obtiene el resultado.

□

Nuevamente cuando la sub- σ -álgebra \mathcal{G} es $\sigma(Y)$, para alguna variable aleatoria Y , entonces $\text{Var}(X|\mathcal{G})$ se escribe $\text{Var}(X|Y)$, y puede tomarse como definición cualquiera de las siguientes expresiones:

$$\text{Var}(X|Y) = \begin{cases} E((X - E(X|Y))^2 | Y) \\ E(X^2|Y) - E^2(X|Y). \end{cases}$$

EJERCICIO. Demuestre que la esperanza de la variable $\text{Var}(X|\mathcal{G})$ es:

a) $E(X^2) - E(E^2(X|\mathcal{G}))$. b) $E(X - E(X|\mathcal{G}))^2$. .

4.5. Ejercicios

Esperanza condicional

402. Demuestre que si c es una constante, entonces $E(c | \mathcal{G}) = c$, para cualquier sub- σ -álgebra \mathcal{G} .
403. Sea A un evento. Demuestre que $E(1_A | \{\emptyset, \Omega\}) = P(A)$.
404. Sea X integrable y sea Y constante. Demuestre que $E(X | Y) = E(X)$.
405. Sea X una variable aleatoria con esperanza finita. Demuestre que $E(X | \{\emptyset, \Omega\}) = E(X)$.
406. Sean A y B dos eventos tales que $0 < P(B) < 1$. Demuestre que

$$E(1_A | 1_B)(\omega) = \begin{cases} P(A | B) & \text{si } \omega \in B, \\ P(A | B^c) & \text{si } \omega \notin B. \end{cases}$$

407. Sea X con esperanza finita y sea B un evento tal que $0 < P(B) < 1$. Demuestre que $E(X | \{\emptyset, B, B^c, \Omega\}) = E(X | B) 1_B + E(X | B^c) 1_{B^c}$.
408. Encuentre $E(X | Y)$ cuando X y Y se distribuyen de manera conjunta de acuerdo a la siguiente tabla.

$x \backslash y$	-1	0	1
1	2/12	2/12	2/12
2	3/12	2/12	1/12

409. Encuentre una distribución conjunta de dos variables aleatorias X y Y de tal forma que $E(X | Y)$ tenga distribución $\text{unif}\{-1, 1\}$.
410. Se ha demostrado que si $X \geq 0$ es integrable, entonces $E(X | \mathcal{G}) \geq 0$. Demuestre que el recíproco es en general falso.
411. Sea c una constante. Diga falso o verdadero. Demuestre o proporcione un contraejemplo.

- a) $E(X | X) = X$.
b) $E(X^2 | X) = X^2$.
c) $E(X | X^2) = X$.
- d) $E(X | cX) = X$.
e) $E(X | X + c) = X$.
f) $E(X | X + Y) = X$.

412. Demuestre que $E(E^2(X | \mathcal{G})) = E(X E(X | \mathcal{G}))$.

413. Sea B_1, \dots, B_n una partición finita de Ω en donde cada elemento tiene probabilidad positiva, y sean b_1, \dots, b_n constantes cualesquiera. Defina la variable aleatoria discreta

$$Y = \sum_{i=1}^n b_i 1_{B_i}.$$

Sea X con segundo momento finito. Demuestre que la distancia entre X y Y definida por $d(X, Y) = [E(X - Y)^2]^{1/2}$ es mínima cuando $b_i = E(X | B_i)$, es decir, cuando la variable Y es la esperanza condicional $E(X | Y)$. Sugerencia: observe que $E(X - Y)^2 = \sum_{i=1}^n E[(X - b_i)^2 | B_i] P(B_i)$, y la suma es mínima si, y sólo si, cada sumando lo es.

414. DESIGUALDAD DE CAUCHY-SCHWARZ CONDICIONAL. Sean X y Y con segundo momento finito. Demuestre que

$$E^2(XY | \mathcal{G}) \leq E(X^2 | \mathcal{G}) E(Y^2 | \mathcal{G}).$$

Sugerencia: proceda como en la desigualdad de Cauchy-Schwarz en el caso no condicional, vea el ejercicio 191.

415. DESIGUALDAD DE MARKOV CONDICIONAL. Sea $X \geq 0$ integrable. Demuestre que para cualquier constante $\epsilon > 0$,

$$P(X \geq \epsilon | \mathcal{G}) \leq \frac{1}{\epsilon} E(X | \mathcal{G}).$$

Sugerencia: Vea la demostración de la desigualdad de Markov no condicional en la página 347.

416. Sean X_1, X_2, \dots independientes idénticamente distribuidas y con esperanza finita. Defina $S_n = X_1 + \dots + X_n$. Demuestre que para $1 \leq k \leq n$,

- a) $E(X_k | S_n) = S_n/n$.
 b) $E(S_k | S_n) = k S_n/n$.
 c) $E(X_k | \sigma\{S_n, S_{n+1}, \dots\}) = S_n/n$.
 d) $E(S_k | \sigma\{S_n, S_{n+1}, \dots\}) = k S_n/n$.

Varianza condicional

417. Demuestre que

- a) $\text{Var}(X | \{\emptyset, \Omega\}) = \text{Var}(X)$.
 b) $\text{Var}(1_A | \{\emptyset, \Omega\}) = P(A)(1 - P(A))$.

418. Encuentre $\text{Var}(X | Y)$ cuando X y Y se distribuyen de manera conjunta de acuerdo a la siguiente tabla.

$x \backslash y$	-1	0	1
1	2/12	2/12	2/12
2	3/12	2/12	1/12

419. Demuestre que $E(\text{Var}(X | \mathcal{G}) | \mathcal{G}) = \text{Var}(X | \mathcal{G})$.

420. Demuestre que $\text{Var}(X) \geq \text{Var}(X | \mathcal{G})$.

421. Demuestre que $E(X^2 | \mathcal{G}) \geq E^2(X | \mathcal{G})$.

422. Demuestre que

- a) si X es \mathcal{G} -medible, entonces $\text{Var}(X | \mathcal{G}) = 0$.
 b) $\text{Var}(\text{Var}(X | \mathcal{G}) | \mathcal{G}) = 0$.
 c) si X es \mathcal{G} -medible, entonces $\text{Var}(\text{Var}(X | \mathcal{G})) = 0$.

CAPÍTULO 5

Transformaciones

Sea X una variable aleatoria con distribución conocida, y sea φ es una función tal que $Y = \varphi(X)$ es otra variable aleatoria. ¿Cuál es la distribución de Y ? En este capítulo se da respuesta a esta pregunta tanto en el caso unidimensional como en el caso de vectores aleatorios. En particular, se encuentran fórmulas explícitas para la función de densidad de la suma, diferencia, producto y cociente de dos variables aleatorias absolutamente continuas.

5.1. Transformación de una variable aleatoria

En esta sección se estudian un par de resultados que proveen de fórmulas para la función de densidad de la variable $\varphi(X)$, en términos de la función de densidad de X . Gráficamente tal transformación se muestra en la Figura 5.1.

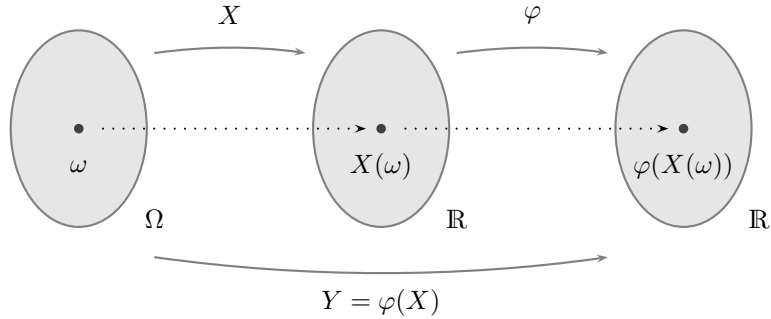


Figura 5.1: La composición $Y = \varphi \circ X$.

TEOREMA DE CAMBIO DE VARIABLE 1. Sea X una variable aleatoria continua con valores dentro de un intervalo $(a, b) \subseteq \mathbb{R}$, y con función de densidad $f_X(x)$. Sea $\varphi : (a, b) \rightarrow \mathbb{R}$ una función continua, estrictamente creciente o decreciente, y con inversa diferenciable. Entonces la variable aleatoria $Y = \varphi(X)$ toma valores dentro del intervalo $\varphi(a, b)$, y tiene función de densidad

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \left| \frac{d}{dy} \varphi^{-1}(y) \right| & \text{para } y \in \varphi(a, b), \\ 0 & \text{otro caso.} \end{cases}$$

Demostración. Suponga primero el caso φ estrictamente creciente. Entonces para $y \in \varphi(a, b)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P(X \leq \varphi^{-1}(y)) \\ &= F_X(\varphi^{-1}(y)). \end{aligned}$$

Derivando se obtiene $f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \frac{d}{dy}\varphi^{-1}(y)$. Para φ estrictamente decreciente,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P(X \geq \varphi^{-1}(y)) \\ &= 1 - F_X(\varphi^{-1}(y)). \end{aligned}$$

Entonces $f_Y(y) = f_X(\varphi^{-1}(y)) \cdot [-\frac{d}{dy}\varphi^{-1}(y)]$. En cualquiera caso se obtiene el resultado enunciado. \square

Por ejemplo, la función $\varphi(x) = e^x$, definida sobre toda la recta real cumple con las condiciones del teorema anterior. Usaremos esta función para mostrar con dos ejemplos la forma de aplicar este resultado.

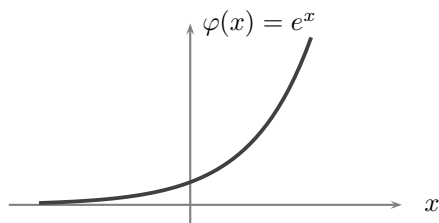


Figura 5.2: La transformación $\varphi(x) = e^x$.

EJEMPLO. (DISTRIBUCIÓN LOG NORMAL). Sea X con distribución $N(\mu, \sigma^2)$, y sea φ la función estrictamente creciente $\varphi(x) = e^x$, con inversa diferenciable $\varphi^{-1}(y) = \ln y$. Entonces la variable aleatoria $Y = e^X$ toma valores en el intervalo $(0, \infty)$, y su distribución se conoce con el nombre de distribución log normal (μ, σ^2) . Su función de densidad tiene la siguiente expresión cuya

gráfica ha sido mostrada antes en la Figura 2.24.

$$f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right] & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

EJEMPLO. (DISTRIBUCIÓN LOG GAMA). Sea X con distribución gama(n, λ), y sea nuevamente $\varphi(x) = e^x$, con inversa diferenciable $\varphi^{-1}(y) = \ln y$. Entonces la variable aleatoria $Y = e^X$ toma valores en el intervalo $(0, \infty)$, y su distribución se conoce como distribución log gama(n, λ). Su función de densidad es

$$f_Y(y) = \begin{cases} \frac{(\lambda \ln y)^{n-1}}{\Gamma(n)} \lambda y^{-\lambda-1} & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

El resultado anterior puede extenderse al caso en el que la transformación φ es estrictamente monótona por pedazos. Se enuncia y demuestra a continuación este resultado cuando la transformación se descompone en dos partes monótonas, siendo fácil la extensión cuando se tiene un mayor número de secciones.

EJERCICIO. Encuentre la función de densidad del cuadrado de una variable aleatoria con distribución $\exp(\lambda)$.

EJERCICIO. Sea X con distribución $\text{unif}(0, 1)$ y sea $\lambda > 0$. Demuestre que

la variable aleatoria $Y = -(1/\lambda) \ln X$ tiene distribución $\exp(\lambda)$. .

TEOREMA DE CAMBIO DE VARIABLE 2. Sea X una variable aleatoria continua con valores dentro de un intervalo $(a, c) \subseteq \mathbb{R}$, y con función de densidad $f_X(x)$. Sea $\varphi : (a, c) \rightarrow \mathbb{R}$ una función tal que admite la descomposición

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{si } x \in (a, b), \\ \varphi_2(x) & \text{si } x \in (b, c), \end{cases}$$

en donde $a < b < c$, y cada una de las funciones $\varphi_1(x) : (a, b) \rightarrow \mathbb{R}$ y $\varphi_2(x) : (b, c) \rightarrow \mathbb{R}$ es continua, estrictamente creciente o decreciente, y con inversa diferenciable. Entonces la variable aleatoria $Y = \varphi(X)$ toma valores dentro del intervalo $\varphi(a, c)$, y tiene función de densidad

$$\begin{aligned} f_Y(y) &= f_X(\varphi_1^{-1}(y)) \left| \frac{d}{dy} \varphi_1^{-1}(y) \right| \cdot 1_{\varphi_1(a,b)}(y) \\ &\quad + f_X(\varphi_2^{-1}(y)) \left| \frac{d}{dy} \varphi_2^{-1}(y) \right| \cdot 1_{\varphi_2(b,c)}(y). \end{aligned}$$

Demostración. La prueba es análoga al caso anterior, únicamente hay que hacer el análisis sobre cada uno de los intervalos de monotonía estricta. Para cualquier y en \mathbb{R} ,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P[(\varphi_1(X) \leq y) \cap (X \in (a, b))] \\ &\quad + P[(\varphi_2(X) \leq y) \cap (X \in (b, c))]. \end{aligned}$$

Nos interesa el comportamiento de estas probabilidades como funciones de y , puesto que calcularemos la derivada de ellas para encontrar $f_Y(y)$. Por ejemplo, la primera probabilidad, vista como función de y , es

$$y \mapsto P[(\varphi_1(X) \leq y) \cap (X \in (a, b))],$$

que permanece constante para $y \notin \varphi_1(a, b)$, de modo que, suponiendo por ejemplo φ_1 creciente, y para $y \in \varphi_1(a, b)$,

$$\begin{aligned} \frac{d}{dy} P[(\varphi_1(X) \leq y) \cap (X \in (a, b))] &= \frac{d}{dy} P[(X \leq \varphi_1^{-1}(y)) \cap (X \in (a, b))] \\ &= \frac{d}{dy} P[a < X \leq \varphi_1^{-1}(y)] \\ &= \frac{d}{dy} F_X(\varphi_1^{-1}(y)) \\ &= f_X(\varphi_1^{-1}(y)) \frac{d}{dy} \varphi_1^{-1}(y). \end{aligned}$$

De manera análoga se procede respecto del segundo sumando, considerando también el caso cuando se presenta la monotonía decreciente. De esta forma se obtiene la fórmula enunciada. \square

EJEMPLO. Sea X continua con función de densidad $f_X(x)$. Considere la transformación $\varphi(x) = x^2$, la cual es estrictamente decreciente en $(-\infty, 0)$, y estrictamente creciente en $(0, \infty)$.

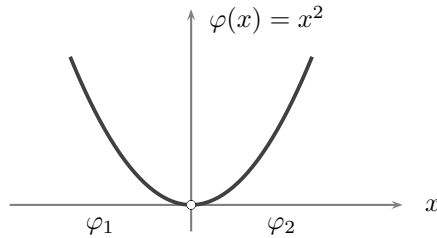


Figura 5.3: La transformación $\varphi(x) = x^2$ como dos secciones monótonas.

Defina entonces las funciones monótonas $\varphi_1(x) = x^2$ sobre $(-\infty, 0)$, y $\varphi_2(x) = x^2$ sobre $(0, \infty)$. Entonces sus inversas son $\varphi_1^{-1}(y) = -\sqrt{y}$, y $\varphi_2^{-1}(y) = \sqrt{y}$. La variable $Y = X^2$ tiene por lo tanto función de densidad

$$f_Y(y) = \begin{cases} f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

EJERCICIO. Sea X con distribución uniforme en el intervalo $(0, 1)$. Demuestre que la función de densidad de la variable $Y = 4X(1 - X)$ es

$$f(y) = \begin{cases} \frac{1}{2\sqrt{1-y}} & \text{si } 0 < y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

EJERCICIO. Sea X con distribución normal estándar. Demuestre que la variable aleatoria $Y = |X|$ tiene función de densidad

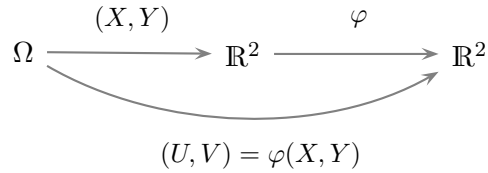
$$f(y) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-y^2/2} & \text{si } y > 0, \\ 0 & \text{si } y \leq 0. \end{cases}$$

EJERCICIO. Sea X con distribución uniforme en el intervalo $[-1, 1]$. Encuentre la función de densidad de la variable $Y = \sqrt{1 - X^2}$.

5.2. Transformación de un vector aleatorio

Suponga ahora que (X, Y) es un vector con función de densidad conocida, y $\varphi(x, y)$ es una función definida en algún subconjunto de \mathbb{R}^2 y con valores en \mathbb{R}^2 . El problema es encontrar la función de densidad del nuevo vector $\varphi(X, Y)$. Gráficamente esta transformación se ilustra en la Figura 5.4.

La transformación $\varphi(x, y)$ se escribirá como $(\varphi_1(x, y), \varphi_2(x, y))$, y la derivada de la primera componente respecto de x , por ejemplo, se escribe $\partial_x \varphi_1$.

Figura 5.4: La composición $\varphi \circ (X, Y)$.

TEOREMA DE CAMBIO DE VARIABLE 3. Sea (X, Y) un vector continuo con valores en $I \subseteq \mathbb{R}^2$, y con función de densidad $f_{X,Y}(x, y)$. Sea $\varphi(x, y) : I \rightarrow \mathbb{R}^2$ una función continua con inversa $\varphi^{-1}(u, v)$ diferenciable. Entonces el vector $(U, V) = \varphi(X, Y)$ toma valores en $\varphi(I)$ y tiene función de densidad

$$f_{U,V}(u, v) = \begin{cases} f_{X,Y}(\varphi^{-1}(u, v)) |J(u, v)| & \text{para } (u, v) \in \varphi(I), \\ 0 & \text{otro caso,} \end{cases} \quad (5.1)$$

en donde

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix}.$$

Una prueba rigurosa de este teorema resulta ser un tanto elaborada y la omitiremos. Sin embargo, puede usarse el siguiente argumento intuitivo para encontrar la fórmula enunciada. Sea

$$(U, V) = \varphi(X, Y) = (\varphi_1(X, Y), \varphi_2(X, Y)),$$

con inversa

$$(X, Y) = \varphi^{-1}(U, V) = (\varphi_1^{-1}(U, V), \varphi_2^{-1}(U, V)).$$

Sea A el rectángulo de área infinitesimal de esquinas con coordenadas (x, y) , $(x + dx, y)$, $(x, y + dy)$ y $(x + dx, y + dy)$. Bajo la transformación φ las coordenadas de las esquinas del rectángulo A se transforman en las siguientes

coordenadas:

$$(x, y) \mapsto (\varphi_1(x, y), \varphi_2(x, y)).$$

$$\begin{aligned} (x + dx, y) &\mapsto (\varphi_1(x + dx, y), \varphi_2(x + dx, y)) \\ &\doteq (\varphi_1(x, y) + \partial_x \varphi_1(x, y)dx, \varphi_2(x, y) \\ &\quad + \partial_x \varphi_2(x, y)dx). \end{aligned}$$

$$\begin{aligned} (x, y + dy) &\mapsto (\varphi_1(x, y + dy), \varphi_2(x, y + dy)) \\ &\doteq (\varphi_1(x, y) + \partial_y \varphi_1(x, y)dy, \varphi_2(x, y) \\ &\quad + \partial_y \varphi_2(x, y)dy). \end{aligned}$$

$$\begin{aligned} (x + dx, y + dy) &\mapsto (\varphi_1(x + dx, y + dy), \varphi_2(x + dx, y + dy)) \\ &\doteq (\varphi_1(x, y) + \partial_x \varphi_1(x, y)dx + \partial_y \varphi_1(x, y)dy, \\ &\quad \varphi_2(x, y) + \partial_x \varphi_2(x, y)dx + \partial_y \varphi_2(x, y)dy). \end{aligned}$$

Gráficamente la transformación de estos puntos se muestra en la Figura 5.5.

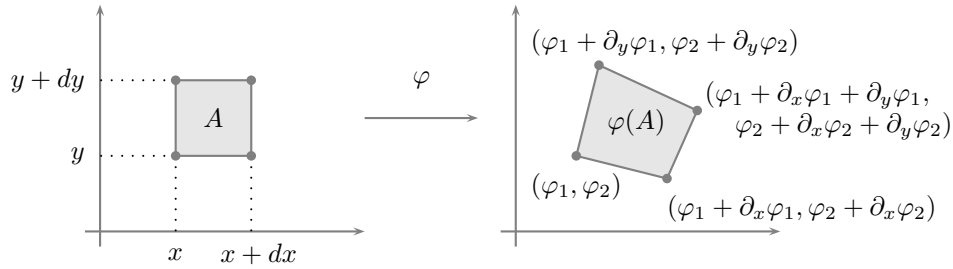


Figura 5.5: La transformación φ aplicada al rectángulo A .

Entonces $P((X, Y) \in A) = P((U, V) \in \varphi(A))$. Por lo tanto

$$f_{X,Y}(x, y) dx dy = f_{U,V}(u, v) \times \text{“Área de } \varphi(A)\text{”}.$$

En donde

$$\begin{aligned}
 \text{“Área de } \varphi(A)\text{”} &= |\partial_x \varphi_1 \cdot \partial_y \varphi_2 - \partial_x \varphi_2 \cdot \partial_y \varphi_1| \, dx dy \\
 &= \left| \begin{vmatrix} \partial_x \varphi_1 & \partial_y \varphi_1 \\ \partial_x \varphi_2 & \partial_y \varphi_2 \end{vmatrix} \right| dx dy \\
 &= |J(x, y)| \, dx dy.
 \end{aligned}$$

Además $|J(x, y)| = \frac{1}{|J(u, v)|}$. Por lo tanto

$$f_{X,Y}(x, y) \, dx dy = f_{U,V}(u, v) \frac{dx dy}{|J(u, v)|}.$$

Es decir, $f_{U,V}(u, v) = f_{X,Y}(\varphi_1^{-1}(u, v), \varphi_2^{-1}(u, v)) |J(u, v)|$.

Como ejemplo de aplicación de esta fórmula, en las secciones siguientes encontraremos expresiones para la función de densidad de la suma, diferencia, producto y cociente de dos variables aleatorias.

Las fórmulas generales sobre transformaciones encontradas hasta ahora se resumen en la siguiente tabla, que sólo sirve como referencia general pues no se mencionan las condiciones precisas de su validez.

TRANSFORMACIONES

$$Y = \varphi(X) \Rightarrow f_Y(y) = f_X(\varphi^{-1}(y)) \left| \frac{d}{dy} \varphi^{-1}(y) \right|.$$

$$(U, V) = \varphi(X, Y) \Rightarrow f_{U,V}(u, v) = f_{X,Y}(\varphi^{-1}(u, v)) |J(u, v)|,$$

$$\text{en donde } J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix}.$$

Distribución de la suma

El siguiente resultado proporciona una fórmula para la función de densidad de la suma de dos variables aleatorias absolutamente continuas.

PROPOSICIÓN. Sea (X, Y) un vector absolutamente continuo con función de densidad $f_{X,Y}(x, y)$. Entonces $X + Y$ tiene función de densidad

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u-v, v) dv. \quad (5.2)$$

Demostración. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x + y, y)$, con inversa $\varphi^{-1}(u, v) = (u - v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$$

Por la fórmula (5.1), $f_{X+Y,Y}(u, v) = f_{X,Y}(u - v, v)$. Integrando respecto a v se obtiene (5.2). \square

Observe que haciendo el cambio de variable $z(v) = u - v$ en (5.2) se obtiene la expresión equivalente

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(z, u - z) dz. \quad (5.3)$$

Ello refleja el hecho de que la suma de dos variables aleatorias es conmutativa. En particular, cuando X y Y son independientes, la fórmula (5.2) se reduce a

$$\begin{aligned} f_{X+Y}(u) &= \int_{-\infty}^{\infty} f_X(u - v) f_Y(v) dv \\ &= \int_{-\infty}^{\infty} f_X(u - v) dF_Y(v). \end{aligned} \quad (5.4)$$

Integrando respecto de u e intercambiando el orden de las integrales se obtiene la correspondiente función de distribución

$$F_{X+Y}(u) = \int_{-\infty}^{\infty} F_X(u-v) dF_Y(v).$$

Más generalmente, puede demostrarse que esta fórmula es válida para cualesquiera dos variables aleatorias independientes X y Y , incluyendo el caso cuando una de ellas es discreta y la otra continua.

En el caso cuando X y Y son discretas, independientes y con valores enteros, es sencillo verificar que la función de probabilidad de $X+Y$ es, en completa analogía con (5.4), $f_{X+Y}(u) = \sum_k f_X(u-k)f_Y(k)$, en donde la suma se toma sobre todos los posibles valores enteros k que la variable aleatoria Y puede tomar.

EJERCICIO. Use la fórmula anterior para demostrar que si X y Y son independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$ respectivamente, entonces $X+Y$ tiene distribución $\text{bin}(n+m, p)$. \blacksquare

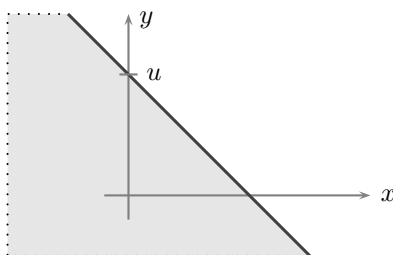
Puede obtenerse la misma fórmula (5.2) mediante el procedimiento usual de encontrar primero la función de distribución de $X+Y$ y después derivar para encontrar la función de densidad. El procedimiento se muestra a continuación.

$$\begin{aligned} F_{X+Y}(u) &= P(X+Y \leq u) \\ &= \int \int_{x+y \leq u} f_{X,Y}(x,y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{u-x} f_{X,Y}(x,y) dy dx. \end{aligned}$$

La región de integración se muestra en la Figura 5.6.

Derivando respecto a u se obtiene

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u-x) dx,$$

Figura 5.6: Región de integración $x + y \leq u$.

que corresponde a la expresión (5.3) equivalente a (5.2).

EJERCICIO. Sean X y Y independientes cada una con distribución normal estándar. Use (5.2) para demostrar que $X + Y$ tiene distribución $N(0, 2)$, es decir, su función de densidad es

$$f(u) = \frac{1}{2\sqrt{\pi}} e^{-u^2/4}.$$

▪

EJERCICIO. Encuentre la función de densidad de $X + Y$ cuando X y Y tienen función de densidad conjunta

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

▪

EJERCICIO. Sea (X, Y, Z) un vector absolutamente continuo con función de densidad $f_{X,Y,Z}(x, y, z)$. Demuestre que la variable $X + Y + Z$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(u - y - z, y, z) dy dz.$$

Aplique esta fórmula para encontrar la función de densidad de la suma de tres variables aleatorias independientes, en donde cada sumando tiene

distribución $\text{unif}(0, 1)$. ▪

CONVOLUCIÓN. La convolución de dos funciones de densidad continuas f_1 y f_2 , es una función de densidad denotada por $f_1 * f_2$, y definida como sigue

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(x - y)f_2(y) dy.$$

Más generalmente, la convolución de dos funciones de distribución F_1 y F_2 es la función de distribución

$$(F_1 * F_2)(x) = \int_{-\infty}^{\infty} F_1(x - y)dF_2(y).$$

En consecuencia, si X y Y son dos variables aleatorias independientes con correspondientes funciones de distribución F_X y F_Y , entonces la función de distribución de la variable $X + Y$ es la convolución $F_X * F_Y$. No es difícil comprobar que $F_X * F_Y = F_Y * F_X$. En particular, la suma de n variables aleatorias independientes todas con la misma función de distribución F tiene función de distribución $F * \cdots * F$, que se escribe simplemente como F^{*n} .

Observe que hemos denotado la convolución por el mismo símbolo, primero cuando los argumentos son funciones de densidad y en el otro cuando son funciones de distribución. Para el caso de funciones de distribución absolutamente continuas, se tiene la relación

$$\frac{d}{dx}(F_1 * F_2)(x) = (f_1 * f_2)(x).$$

Distribución de la diferencia

Se encontrará ahora una fórmula para la función de densidad de la diferencia de dos variables aleatorias.

PROPOSICIÓN. Sea (X, Y) un vector absolutamente continuo con función de densidad $f_{X,Y}(x, y)$. Entonces $X - Y$ tiene función de densidad

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u+v, v) dv. \quad (5.5)$$

Demostración. Procedemos como en la sección anterior. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x - y, y)$ con inversa $\varphi^{-1}(u, v) = (u + v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

Por la fórmula (5.1), $f_{X-Y}(u, v) = f_{X,Y}(u + v, v)$. Integrando respecto a v se obtiene (5.5). \square

Con el cambio de variable $z(v) = u + v$ en (5.5) se obtiene la expresión equivalente

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(z, z - u) dz. \quad (5.6)$$

Cuando X y Y son independientes la fórmula (5.5) se reduce a

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_X(u + v) f_Y(v) dv.$$

En el caso discreto cuando X y Y son independientes con valores enteros, la variable $X - Y$ también toma valores enteros, y tiene función de probabilidad

$$f_{X-Y}(u) = \sum_k f_X(u + k) f_Y(k),$$

en donde la suma se toma sobre todos los posibles valores enteros k que Y puede tomar.

Nuevamente se puede demostrar (5.5) mediante el procedimiento usual de encontrar primero la función de distribución y después derivar para encontrar la función de densidad. Por definición,

$$\begin{aligned} F_{X-Y}(u) &= P(X - Y \leq u) \\ &= \int \int_{x-y \leq u} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{x-u}^{\infty} f_{X,Y}(x, y) dy dx. \end{aligned}$$

La región de integración aparece en la Figura 5.7.

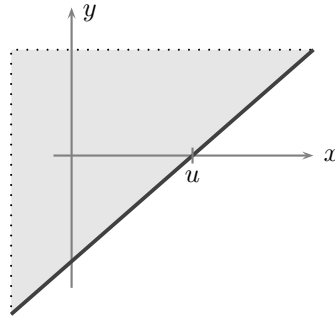


Figura 5.7: Región de integración $x - y \leq u$.

Derivando respecto a u se obtiene (5.6) equivalente a (5.5). A partir de la fórmula para la suma de dos variables aleatorias se puede construir una tercera demostración de (5.5). Por la fórmula para la suma,

$$f_{X-Y}(u) = f_{X+(-Y)}(u) = \int_{-\infty}^{\infty} f_{X,-Y}(u - v, v) dv.$$

Haciendo el cambio de variable $x = -v$, se obtiene

$$\begin{aligned} f_{X-Y}(u) &= \int_{-\infty}^{\infty} f_{X,-Y}(u + x, -x) dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(u + x, x) dx. \end{aligned}$$

EJERCICIO. Sean X y Y independientes cada una con distribución normal estándar. Use (5.5) para demostrar que $X - Y$ tiene distribución $N(0, 2)$, es decir, su función de densidad es

$$f(u) = \frac{1}{2\sqrt{\pi}}e^{-u^2/4}.$$

▪

EJERCICIO. Encuentre la función de densidad de $Y - X$ cuando X y Y tienen función de densidad conjunta

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

▪

EJERCICIO. Sean X y Y independientes cada una con distribución normal estándar. En ejercicios anteriores se ha pedido comprobar que tanto $X + Y$ como $X - Y$ tienen distribución $N(0, 2)$. Demuestre ahora que $X + Y$ y $X - Y$ son independientes.

▪

EJERCICIO. Sea (X, Y, Z) un vector absolutamente continuo con función de densidad $f_{X,Y,Z}(x, y, z)$. Demuestre que la variable $X - Y - Z$ tiene función de densidad

$$f_{X-Y-Z}(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(u + y + z, y, z) dy dz.$$

Aplique esta fórmula para encontrar la función de densidad de $X - Y - Z$, cuando estas variables son independientes y cada una de ellas tiene distribución $\text{unif}(0, 1)$.

▪

Distribución del producto

Ahora se encontrará una fórmula para la función de densidad del producto de dos variables aleatorias absolutamente continuas.

PROPOSICIÓN. Sea (X, Y) un vector absolutamente continuo con función de densidad $f_{X,Y}(x, y)$. Entonces XY tiene función de densidad

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u/v, v) |1/v| dv. \quad (5.7)$$

Demostración. Se usa nuevamente la fórmula (5.1). Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (xy, y)$ cuya inversa es, para $v \neq 0$, $\varphi^{-1}(u, v) = (u/v, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} 1/v & u/v^2 \\ 0 & 1 \end{vmatrix} = 1/v.$$

Por la fórmula (5.1), para $v \neq 0$, $f_{XY}(u, v) = f_{X,Y}(u/v, v) |1/v|$. Integrando respecto a v se obtiene (5.7). \square

Haciendo $x(v) = u/v$ en (5.7) se obtiene la expresión equivalente

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u/x) |1/x| dx. \quad (5.8)$$

Cuando X y Y son independientes (5.7) se reduce a

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_X(u/v) f_Y(v) |1/v| dv.$$

Usaremos nuevamente el procedimiento usual de encontrar primero la función de distribución de XY y después derivar para encontrar la función de densidad. Por definición,

$$\begin{aligned} F_{XY}(u) &= P(XY \leq u) \\ &= \int \int_{xy \leq u} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^0 \int_{u/x}^{\infty} f_{X,Y}(x, y) dy dx + \int_0^{\infty} \int_{-\infty}^{u/x} f_{X,Y}(x, y) dy dx. \end{aligned}$$

La región de integración se muestra en la Figura 5.8.

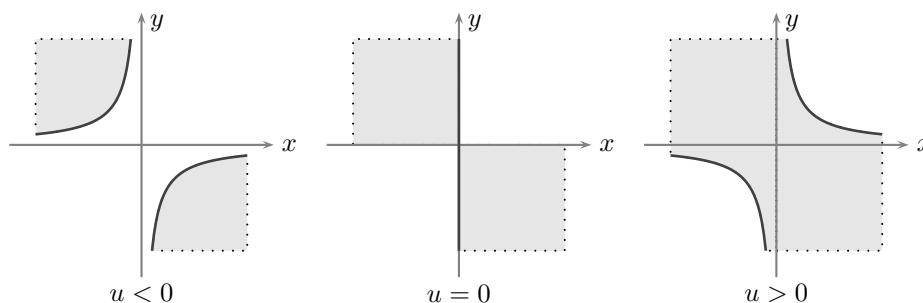


Figura 5.8: Región de integración $xy \leq u$.

Derivando respecto a u ,

$$\begin{aligned} f_{XY}(u) &= \int_{-\infty}^0 f_{X,Y}(x, u/x)(-1/x) dy dx \\ &\quad + \int_0^{\infty} f_{X,Y}(x, u/x)(1/x) dy dx. \\ &= \int_{-\infty}^{\infty} f_{X,Y}(x, u/x)|1/x| dx, \end{aligned}$$

que corresponde a (5.8), equivalente a (5.7).

EJERCICIO. Encuentre la función de densidad de XY cuando X y Y tienen función de densidad conjunta

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

▪

EJERCICIO. Sea (X, Y, Z) un vector absolutamente continuo con función de densidad $f_{X,Y,Z}(x, y, z)$. Demuestre que la variable XYZ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}\left(\frac{u}{vw}, v, w\right) \left|\frac{1}{vw}\right| dv dw.$$

Aplique este resultado al caso cuando X , Y y Z son independientes cada una de ellas con distribución $\text{unif}(0, 1)$. \square

Distribución del cociente

Finalmente se encontrará una fórmula para el cociente de dos variables aleatorias absolutamente continuas.

PROPOSICIÓN. Sea (X, Y) un vector absolutamente continuo con función de densidad $f_{X,Y}(x, y)$ y tal que $Y \neq 0$. Entonces X/Y tiene función de densidad

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(uv, v) |v| dv. \quad (5.9)$$

Demostración. Procederemos como en las secciones anteriores. Sea $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la transformación $\varphi(x, y) = (x/y, y)$ para $y \neq 0$, y con inversa $\varphi^{-1}(u, v) = (uv, v)$. El Jacobiano de la transformación inversa es

$$J(u, v) = \begin{vmatrix} \partial_u \varphi_1^{-1} & \partial_v \varphi_1^{-1} \\ \partial_u \varphi_2^{-1} & \partial_v \varphi_2^{-1} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v.$$

Por la fórmula (5.1), $f_{X/Y,Y}(u, v) = f_{X,Y}(uv, v) |v|$, de donde se obtiene (5.9) integrando respecto a v . \square

Haciendo $x(v) = uv$ en (5.9) se obtiene la expresión equivalente

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, x/u) |x/u^2| dx. \quad (5.10)$$

Observe nuevamente que cuando X y Y son independientes, el integrando en la fórmula (5.9) se escribe como el producto de las densidades marginales.

Ahora usaremos el procedimiento usual de encontrar primero la función de distribución y después derivar para encontrar la función de densidad.

$$\begin{aligned}
 F_{X/Y}(u) &= P(X/Y \leq u) \\
 &= \int \int_{x/y \leq u} f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^0 \int_{uy}^{\infty} f_{X,Y}(x, y) dx dy + \int_0^{\infty} \int_{-\infty}^{uy} f_{X,Y}(x, y) dx dy.
 \end{aligned}$$

La región de integración se muestra en la Figura 5.9.

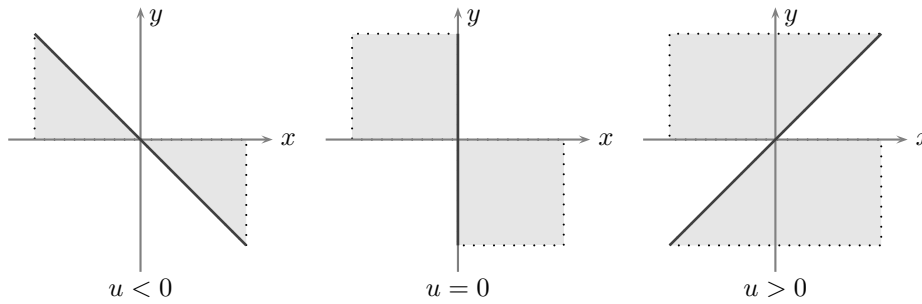


Figura 5.9: Región de integración $x/y \leq u$.

Derivando respecto a u ,

$$\begin{aligned}
 f_{X/Y}(u) &= - \int_{-\infty}^0 f_{X,Y}(uy, y) y dy + \int_0^{\infty} f_{X,Y}(uy, y) y dy \\
 &= \int_{-\infty}^{\infty} f_{X,Y}(uy, y) |y| dy.
 \end{aligned}$$

A partir de la fórmula para el producto de dos variables aleatorias se puede construir una tercera demostración de (5.9) de la siguiente forma.

$$f_{X/Y}(u) = f_{X \cdot (1/Y)}(u) = \int_{-\infty}^{\infty} f_{X,1/Y}(u/v, v) |1/v| dv.$$

Haciendo el cambio de variable $x = 1/v$ se obtiene

$$\begin{aligned} f_{X/Y}(u) &= \int_{-\infty}^{\infty} f_{X,1/Y}(ux, 1/x)|x| dx \\ &= \int_{-\infty}^{\infty} f_{X,Y}(ux, x)|x| dx. \end{aligned}$$

EJERCICIO. Sean X y Y independientes con distribución normal estándar. Demuestre que X/Y tiene distribución Cauchy, es decir, su función de densidad es

$$f(u) = \frac{1}{\pi(1+u^2)}, \quad \text{para } -\infty < u < \infty.$$

▪

EJERCICIO. Encuentre la función de densidad de X/Y cuando X y Y tienen función de densidad conjunta

$$f(x, y) = \begin{cases} 3(x^2 + y^2)/16 & \text{si } 0 < x < y < 2, \\ 0 & \text{otro caso.} \end{cases}$$

▪

EJERCICIO. Sea (X, Y, Z) un vector absolutamente continuo con función de densidad $f_{X,Y,Z}(x, y, z)$. Demuestre que la variable $X/(YZ)$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(uvw, v, w) |vw| dv dw.$$

Aplique este resultado al caso cuando X , Y y Z son independientes cada una de ellas con distribución $\text{unif}(0, 1)$.

▪

Las fórmulas para la función de densidad de las operaciones básicas entre dos variables aleatorias encontradas en este capítulo se resumen en la siguiente tabla.

FÓRMULAS PARA LA SUMA, DIFERENCIA, PRODUCTO Y COCIENTE
DE DOS VARIABLES ALEATORIAS ABSOLUTAMENTE CONTINUAS

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u-v, v) dv$$

$$f_{X-Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u+v, v) dv$$

$$f_{XY}(u) = \int_{-\infty}^{\infty} f_{X,Y}(u/v, v) |1/v| dv$$

$$f_{X/Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(uv, v) |v| dv$$

5.3. Ejercicios

Transformación de una variable aleatoria

423. Sea X con distribución $\exp(\lambda)$. Encuentre la función de densidad y de distribución de la variable $Y = 1 - \exp(-\lambda X)$.
424. Encuentre la distribución de $Y = 1/X$ cuando X tiene distribución:
a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.
425. Sea X continua con función de densidad $f_X(x)$. Demuestre que

$$f_{|X|}(x) = \begin{cases} f_X(-x) + f_X(x) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

426. Sea X con distribución uniforme en el intervalo $(0, 2\pi)$. Encuentre la función de densidad de la variable
- a) $Y = \text{sen}(X)$.
b) $Y = \text{cos}(X)$.
427. Encuentre la distribución de $Y = X^n$ para cada n en \mathbb{N} , cuando X tiene distribución
- a) $\text{unif}(0, 1)$.
b) $\text{unif}(-1, 1)$.
c) $\exp(\lambda)$.
428. Sea X con distribución $\text{unif}(-1, 1)$. Encuentre la función de densidad de X^2 .
429. Sea X absolutamente continua con función de distribución $F(x)$. Demuestre que $Y = F(X)$ tiene distribución $\text{unif}[0, 1]$.

430. Encuentre la función de densidad de $Y = 1/X$ cuando X tiene función de densidad

$$f_X(x) = \begin{cases} 1/2 & \text{si } 0 < x \leq 1, \\ 1/(2x^2) & \text{si } x > 1, \\ 0 & \text{si } x \leq 0. \end{cases}$$

431. Sea X con distribución $\text{unif}(a, b)$. Encuentre la distribución de la variable aleatoria $Y = X/(b - X)$.

Transformación de un vector aleatorio

432. Sean X y Y independientes ambas con distribución $\text{unif}(0, 1)$. Encuentre la función de densidad del vector

- a) $(X, X + Y)$.
b) $(X + Y, X - Y)$.

433. Sean X y Y independientes ambas con distribución $\text{unif}(-1, 1)$. Encuentre la función de densidad del vector

- a) $(X + Y, X - Y)$.
b) $(X, |Y - X|)$.
c) $(X - Y, Y - X)$.

434. Sea (X, Y) un vector con distribución uniforme en el círculo unitario $\{(x, y) : x^2 + y^2 \leq 1\}$. Encuentre la función de densidad del vector

$$(R, \Theta) = (\sqrt{X^2 + Y^2}, \arctan(Y/X)).$$

435. Sean X y Y independientes cada una con distribución $\exp(\lambda)$. Demuestre que el vector $(X, X + Y)$ tiene función de densidad

$$f(u, v) = \begin{cases} \lambda^2 e^{-\lambda v} & \text{si } 0 < u < v, \\ 0 & \text{otro caso.} \end{cases}$$

436. Sea (X, Y) con función de densidad $f_{X,Y}(x, y)$. Demuestre que la función de densidad del vector $(U, V) = (X + Y, X/(X + Y))$ es $f_{U,V}(u, v) = f_{X,Y}(uv, u(1 - v))u$.

Distribución de la suma

437. Encuentre la función de densidad de la suma de dos variables aleatorias cuya función de densidad conjunta es

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

c) $f(x, y) = e^{-y}$, para $0 < x < y$.

d) $f(x, y) = 8xy$, para $0 < x < y < 1$.

e) $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

438. Encuentre la función de densidad de la suma de dos variables aleatorias independientes cada una de ellas con distribución: a) $\text{unif}(0, 1)$.
b) $\exp(\lambda)$.

439. Encuentre la función de densidad de la suma de dos variables aleatorias independientes cada una de ellas con función de densidad

a) $f(x) = 2x$, para $0 < x < 1$.

b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.

440. Encuentre la función de densidad de la suma de dos variables aleatorias independientes X y Y , tales que

a) X tiene distribución $\text{unif}(-1, 0)$ y Y tiene distribución $\text{unif}(0, 1)$.

b) X tiene distribución $\text{unif}(0, 1)$ y Y tiene distribución $\exp(\lambda)$.

441. Encuentre la función de densidad de la suma de dos variables aleatorias con distribución conjunta uniforme en el cuadrado $(-1, 1) \times (-1, 1)$.

442. Encuentre la función de densidad de la suma de tres variables aleatorias con distribución conjunta uniforme en el cubo $(-1, 1) \times (-1, 1) \times (-1, 1)$.

443. Encuentre la función de densidad de la suma de n variables aleatorias con distribución conjunta uniforme en el hipercubo

$$\underbrace{(-1, 1) \times \cdots \times (-1, 1)}_n.$$

444. Demuestre que la suma de dos variables aleatorias independientes, cada una de ellas con distribución normal, tiene nuevamente distribución normal, con media la suma de las medias, y varianza la suma de las varianzas.
445. Sean X_1, \dots, X_n independientes en donde X_k tiene distribución $N(\mu_k, \sigma_k^2)$ para $k = 1, \dots, n$. Sean c_1, \dots, c_n constantes dadas, no todas cero. Demuestre que

$$\sum_{k=1}^n c_k X_k \sim N\left(\sum_{k=1}^n c_k \mu_k, \sum_{k=1}^n c_k^2 \sigma_k^2\right).$$

446. Sean X_1, \dots, X_n independientes y con idéntica distribución $N(\mu, \sigma^2)$. Demuestre que la media muestral dada por $(X_1 + \cdots + X_n)/n$ tiene distribución $N(\mu, \sigma^2/n)$.
447. Demuestre que la suma de dos variables aleatorias independientes, cada una de ellas con distribución $\exp(\lambda)$, tiene distribución $\text{gama}(2, \lambda)$. Más generalmente, demuestre que la suma de n variables aleatorias independientes, cada una de ellas con distribución $\exp(\lambda)$, tiene distribución $\text{gama}(n, \lambda)$.
448. Demuestre que la suma de dos variables aleatorias independientes con distribución $\text{gama}(n, \lambda)$ y $\text{gama}(m, \lambda)$, tiene distribución $\text{gama}(n + m, \lambda)$.
449. Sean X y Y son discretas, independientes y con valores enteros. Demuestre que $f_{X+Y}(u) = \sum_k f_X(u - k)f_Y(k)$, en donde la suma se efectúa sobre todos los posibles valores enteros k que la variable aleatoria Y puede tomar.

450. Sea (X_1, \dots, X_n) un vector aleatorio absolutamente continuo. Demuestre que la variable $X_1 + \dots + X_n$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(u - v_2 - \cdots - v_n, v_2, \dots, v_n) dv_2 \cdots dv_n.$$

Aplique esta fórmula para encontrar la función de densidad de la suma de n variables aleatorias independientes, en donde cada sumando tiene distribución $\text{unif}(0, 1)$.

Distribución de la diferencia

451. Encuentre la función de densidad de $X - Y$, para (X, Y) un vector con función de densidad

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

c) $f(x, y) = e^{-y}$, para $0 < x < y$.

d) $f(x, y) = 8xy$, para $0 < x < y < 1$.

e) $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

452. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y ambas con distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.

453. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y ambas con función de densidad

a) $f(x) = 2x$, para $0 < x < 1$.

b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.

454. Encuentre la función de densidad de $X - Y$, cuando X y Y son independientes y tales que

a) X tiene distribución $\text{unif}(1, 2)$ y Y tiene distribución $\text{unif}(0, 1)$.

b) X tiene distribución $\exp(\lambda)$ y Y tiene distribución $\text{unif}(0, 1)$.

455. Sea a una constante. Demuestre que la diferencia de dos variables aleatorias independientes ambas con distribución uniforme en el intervalo $(a - 1/2, a + 1/2)$ tiene función de densidad

$$f(u) = \begin{cases} 1 - |u| & \text{si } -1 < u < 1, \\ 0 & \text{otro caso.} \end{cases}$$

456. Demuestre que la diferencia de dos variables aleatorias independientes, cada una de ellas con distribución normal, tiene nuevamente distribución normal, con media la diferencia de las medias, y varianzas la *suma* de las varianzas.
457. Sean X y Y son discretas, independientes y con valores enteros. Demuestre que $f_{X-Y}(u) = \sum_k f_X(u+k)f_Y(k)$, en donde la suma se efectúa sobre todos los posibles valores enteros k que la variable aleatoria Y puede tomar.
458. Sea (X, Y, Z) un vector aleatorio absolutamente continuo. Encuentre una fórmula para la función de densidad de la variable $X + Y - Z$.
459. Sea (X, Y, Z) un vector aleatorio absolutamente continuo. Encuentre una fórmula para la función de densidad de la variable $X - Y + Z$.
460. Sea (X_1, \dots, X_n) un vector absolutamente continuo con función de densidad $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$. Demuestre que la variable $X_1 - X_2 - \dots - X_n$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(u + v_2 + \cdots + v_n, v_2, \dots, v_n) dv_2 \cdots dv_n.$$

Aplique esta fórmula al caso cuando las variables X_1, \dots, X_n son independientes y cada una de ellas tiene distribución $\text{unif}(0, 1)$.

Distribución del producto

461. Encuentre la función de densidad del producto de dos variables aleatorias independientes ambas con distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.

462. Encuentre la función de densidad del producto de dos variables aleatorias cuya función de densidad conjunta es

a) $f(x, y) = \frac{1}{ab}$, para $0 < x < a$, $0 < y < b$.

b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

c) $f(x, y) = e^{-y}$, para $0 < x < y$.

d) $f(x, y) = 8xy$, para $0 < x < y < 1$.

e) $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

463. Encuentre la función de densidad del producto de dos variables aleatorias independientes cada una de ellas con función de densidad

a) $f(x) = 2x$, para $0 < x < 1$.

b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.

464. Encuentre la función de densidad del producto de dos variables aleatorias independientes X y Y , tales que

a) X tiene distribución $\text{unif}(-1, 0)$ y Y tiene distribución $\text{unif}(0, 1)$.

b) X tiene distribución $\text{unif}(0, 1)$ y Y tiene distribución $\exp(\lambda)$.

465. Sea (X_1, \dots, X_n) un vector absolutamente continuo con función de densidad $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$. Demuestre que la variable producto $X_1 \cdots X_n$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}\left(\frac{u}{v_2 \cdots v_n}, v_2, \dots, v_n\right) \left| \frac{1}{v_2 \cdots v_n} \right| dv_2 \cdots dv_n.$$

Aplice este resultado al caso cuando todas las variables son independientes y cada una de ellas tiene distribución $\text{unif}(0, 1)$.

Distribución del cociente

466. Encuentre la función de densidad de X/Y para (X, Y) un vector con función de densidad

a) $f(x, y) = \frac{1}{ab}$ para $0 < x < a, 0 < y < b$.

b) $f(x, y) = e^{-x-y}$, para $x, y > 0$.

c) $f(x, y) = e^{-y}$, para $0 < x < y$.

d) $f(x, y) = 8xy$, para $0 < x < y < 1$.

e) $f(x, y) = 4x(1 - y)$, para $0 < x, y < 1$.

f) $f(x, y) = 2e^{-x-y}$, para $0 < x < y$.

467. Encuentre la función de densidad de X/Y cuando X y Y son independientes y ambas con distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.

468. Encuentre la función de densidad de X/Y cuando X y Y son independientes y ambas con función de densidad

a) $f(x) = 2x$, para $0 < x < 1$.

b) $f(x) = 6x(1 - x)$, para $0 < x < 1$.

c) $f(x) = (1 + x)/2$, para $-1 < x < 1$.

469. Encuentre la función de densidad de X/Y cuando X y Y son independientes y son tales que

a) X tiene distribución $\text{unif}(-1, 0)$ y Y tiene distribución $\text{unif}(0, 1)$.

b) X tiene distribución $\text{unif}(0, 1)$ y Y tiene distribución $\exp(\lambda)$.

470. Sean X y Y independientes con distribución $\exp(\lambda)$. Encuentre la función de densidad de $X/(X + Y)$.

471. Sea (X_1, \dots, X_n) un vector absolutamente continuo con función de densidad $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$. Demuestre que la variable $X_1/(X_2 \cdots X_n)$ tiene función de densidad

$$f(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(uv_2 \cdots v_n, v_2, \dots, v_n) |v_2 \cdots v_n| dv_2 \cdots dv_n.$$

Aplique este resultado al caso cuando X , Y y Z son independientes cada una de ellas con distribución $\text{unif}(0, 1)$.

CAPÍTULO 6

Distribuciones muestrales y estadísticas de orden

En este capítulo se estudian algunas distribuciones de probabilidad que surgen en la estadística y otras áreas de aplicación de la probabilidad. Se estudian también algunas fórmulas para las distribuciones de las estadísticas de orden de una muestra aleatoria.

DEFINICIÓN. (MUESTRA ALEATORIA). Una muestra aleatoria es una colección de variables aleatorias X_1, \dots, X_n , que cumplen la condición de ser independientes y de tener cada una de ellas la misma distribución. Al número n se le llama tamaño de la muestra aleatoria.

A menudo se escribe m.a. para abreviar el término *muestra aleatoria*, y se usan las siglas v.a.i.i.d. para denotar el término *variables aleatorias independientes e idénticamente distribuidas*. Por lo tanto, una m.a. es una colección de v.a.i.i.d.

DEFINICIÓN. (ESTADÍSTICA). Una estadística es una variable aleatoria de la forma $g(X_1, \dots, X_n)$, en donde X_1, \dots, X_n es una muestra aleatoria, y $g : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función Borel medible.

EJEMPLO. (MEDIA Y VARIANZA MUESTRAL). La media muestral es una estadística denotada por \bar{X} y definida como sigue

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Observe que \bar{X} es una combinación lineal de los elementos de la m.a. y por lo tanto es una variable aleatoria. Otro ejemplo importante de estadística es la varianza muestral, denotada por S^2 y definida como sigue

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Observe que en el denominador aparece el número de sumandos menos uno. La media y la varianza muestrales tienen la característica de ser estimadores insesgados para la media y la varianza, respectivamente, de una distribución cualquiera. ▪

En particular, cuando la muestra aleatoria proviene de una distribución normal, resulta que la media y la varianza muestrales son independientes. Este es un resultado interesante e inesperado, y la demostración puede encontrarse en [20].

PROPOSICIÓN. Sea X_1, \dots, X_n una m.a. de la distribución $N(\mu, \sigma^2)$. Entonces las estadísticas \bar{X} y S^2 son independientes.

Utilizaremos este resultado más adelante. La proposición recién enunciada no es válida para cualquier distribución de probabilidad, por ejemplo, no es difícil verificar su no validez para una muestra aleatoria de la distribución Bernoulli.

6.1. Distribuciones muestrales

Se estudian a continuación algunas distribuciones que surgen en la estadística al considerar funciones de una muestra aleatoria, en particular, la media y la varianza muestral.

DISTRIBUCIÓN JI-CUADRADA. La variable aleatoria continua X tiene una distribución ji-cuadrada con $n > 0$ grados de libertad, si su función de densidad es

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

En este caso se escribe $X \sim \chi^2(n)$. El término χ^2 se lee ji-cuadrada. La gráfica de esta función de densidad se muestra en la Figura 6.1.

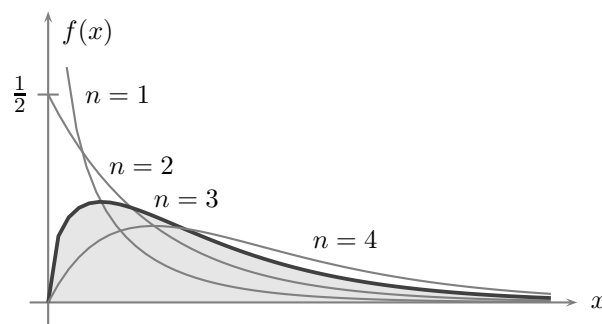


Figura 6.1: Función de densidad $\chi^2(n)$.

Puede demostrarse que $E(X) = n$, y $\text{Var}(X) = 2n$. Observe que la distribución $\chi^2(n)$ con $n = 2$ se reduce a la distribución $\exp(\lambda)$ con $\lambda = 1/2$. La distribución ji-cuadrada puede encontrarse como indican los siguientes resultados.

PROPOSICIÓN. Si $X \sim N(0, 1)$, entonces $X^2 \sim \chi^2(1)$.

Demostración. Para $x > 0$,

$$\begin{aligned}
 f_{X^2}(x) &= f_X(\sqrt{x})\frac{1}{2\sqrt{x}} + f_X(-\sqrt{x})\frac{1}{2\sqrt{x}} \\
 &= f_X(\sqrt{x})\frac{1}{\sqrt{x}} \\
 &= \frac{1}{\sqrt{2\pi}}e^{-x/2}\frac{1}{\sqrt{x}} \\
 &= \frac{1}{\Gamma(1/2)}\left(\frac{1}{2}\right)^{1/2}x^{1/2-1}e^{-x/2}.
 \end{aligned}$$

Esta es la función de densidad de la distribución $\chi^2(1)$. □

La suma de dos o más variables aleatorias independientes con distribución ji-cuadrada es nuevamente una variable aleatoria ji-cuadrada, y sus grados de libertad son la suma de los grados de libertad de cada uno de los sumandos. Este es el contenido de la siguiente proposición.

PROPOSICIÓN. Sean X_1, \dots, X_m independientes tales que cada X_i tiene distribución $\chi^2(n_i)$, para $i = 1, \dots, m$. Entonces

$$\sum_{i=1}^m X_i \sim \chi^2(n_1 + \dots + n_m).$$

Demostración. Es suficiente demostrar el resultado para el caso de dos variables aleatorias. Sean X y Y independientes con distribución ji-cuadrada con grados de libertad n y m , respectivamente. Este ligero cambio en la

notación evitará el uso de subíndices. Por la fórmula (5.2), para $u > 0$,

$$\begin{aligned}
 f_{X+Y}(u) &= \int_0^u f_X(u-v)f_Y(v) dv \\
 &= \int_0^u \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} (u-v)^{n/2-1} e^{-(u-v)/2} \\
 &\quad \frac{1}{\Gamma(m/2)} \left(\frac{1}{2}\right)^{m/2} v^{m/2-1} e^{-v/2} dv \\
 &= \frac{1}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} \\
 &\quad \int_0^u (u-v)^{n/2-1} v^{m/2-1} dv.
 \end{aligned}$$

Haciendo el cambio de variable $w(v) = v/u$ se obtiene

$$\begin{aligned}
 f_{X+Y}(u) &= \frac{1}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1} \\
 &\quad \int_0^1 (1-w)^{n/2-1} w^{m/2-1} dw.
 \end{aligned}$$

La integral resultante es $B(n/2, m/2)$. Entonces

$$\begin{aligned}
 f_{X+Y}(u) &= \frac{B(n/2, m/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1} \\
 &= \frac{1}{\Gamma((n+m)/2)} \left(\frac{1}{2}\right)^{(n+m)/2} e^{-u/2} u^{(n+m)/2-1}.
 \end{aligned}$$

Esta es la función de densidad de la distribución $\chi^2(n+m)$. □

El resultado anterior puede demostrarse de una manera más simple y elegante usando la función generadora de momentos o la función característica, presentadas en el siguiente capítulo.

PROPOSICIÓN. Sean X_1, \dots, X_n independientes cada una con distribución $N(\mu, \sigma^2)$. Entonces

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n).$$

Demostración. Esto es una consecuencia sencilla de las dos proposiciones anteriores. Como cada una de las variables X_i tiene distribución $N(\mu, \sigma^2)$, para $i = 1, \dots, n$, entonces $(X_i - \mu)/\sigma$ tiene distribución $N(0, 1)$. Por lo tanto, $(X_i - \mu)^2/\sigma^2$ tiene distribución $\chi^2(1)$. En consecuencia, $\sum_{i=1}^n (X_i - \mu)^2/\sigma^2$ tiene distribución $\chi^2(n)$. \square

Ahora se enuncia un resultado cuya demostración se pospone hasta que se cuente con la poderosa herramienta de las funciones generadoras de momentos. Este es el contenido del ejercicio 572 en la página 341.

PROPOSICIÓN. Sean X y Y independientes tales que X tiene distribución $\chi^2(n)$, y $X + Y$ tiene distribución $\chi^2(m)$ con $m > n$. Entonces Y tiene distribución $\chi^2(m - n)$.

Con ayuda de esta proposición se demuestra ahora el siguiente resultado de particular importancia en estadística.

PROPOSICIÓN. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. Entonces

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1).$$

Demostración.

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2. \end{aligned}$$

Dividiendo entre σ^2 ,

$$\sum_{i=1}^n \frac{1}{\sigma^2} (X_i - \mu)^2 = \frac{n-1}{\sigma^2} S^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2.$$

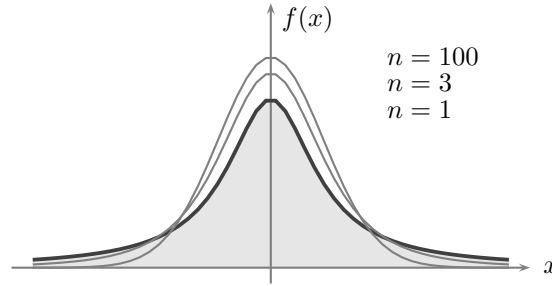
El término del lado izquierdo tiene distribución $\chi^2(n)$, mientras que el segundo sumando del lado derecho tiene distribución $\chi^2(1)$. Por la proposición anterior, y recordando que \bar{X} y S^2 son independientes, se concluye que el primer sumando del lado derecho tiene distribución $\chi^2(n-1)$. \square

DISTRIBUCIÓN t . La variable aleatoria continua X tiene una distribución t de Student con $n > 0$ grados de libertad si su función de densidad está dada por

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} (1 + x^2/n)^{-(n+1)/2}, \quad \text{para } -\infty < x < \infty,$$

cuya gráfica se muestra en la Figura 6.2, cualitativamente es muy parecida a la densidad normal estándar.

En este caso se escribe $X \sim t(n)$. Esta distribución apareció por primera vez en 1908 en un trabajo publicado por William Gosset bajo el pseudónimo de *Student*. Cuando el valor del parámetro n es igual a uno se obtiene la distribución Cauchy. Se puede demostrar también que $E(X) = 0$, y $\text{Var}(X) = n/(n-2)$, para $n > 2$. La primera igualdad establece que esta distribución se encuentra siempre centrada en cero para cualquier valor del parámetro n . Se muestran a continuación algunas formas en las que surge esta distribución.

Figura 6.2: Función de densidad $t(n)$.

PROPOSICIÓN. Sean $X \sim N(0, 1)$ y $Y \sim \chi^2(n)$ independientes. Entonces

$$\frac{X}{\sqrt{Y/n}} \sim t(n).$$

Demostración. Por independencia, la función de densidad conjunta de X y Y es, para $y > 0$,

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} y^{n/2-1} e^{-y/2}.$$

Se aplica la fórmula (5.1) para la transformación $\varphi(x, y) = (x, x/\sqrt{y/n})$, con inversa $\varphi^{-1}(s, t) = (s, ns^2/t^2)$. El Jacobiano de la transformación inversa es

$$J(s, t) = \begin{vmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 2sn/t^2 & -2ns^2/t^3 \end{vmatrix} = -2ns^2/t^3.$$

Por lo tanto

$$\begin{aligned} f_{S,T}(s, t) &= f_X(s) f_Y(ns^2/t^2) \cdot 2ns^2/t^3 \\ &= \frac{1}{\sqrt{2\pi}} e^{-s^2/2} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} \frac{n^{n/2-1} s^{n-2}}{t^{n-2}} e^{-ns^2/2t^2} 2ns^2/t^3. \end{aligned}$$

Integrando respecto a s ,

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{n/2-1}\Gamma(n/2)t^{n+1}} \int_0^\infty s^n e^{-s^2(1+n/t^2)/2} ds.$$

Ahora efectuamos el cambio de variable $r(s) = s^2(1 + n/t^2)/2$, de donde obtenemos $dr = s(1 + n/t^2)ds$, y entonces

$$\begin{aligned} f_T(t) &= \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{n/2-1}\Gamma(n/2)t^{n+1}2\left(\frac{1}{2} + \frac{n}{2t^2}\right)^{(n+1)/2}} \int_0^\infty r^{(n-1)/2} e^{-r} dr \\ &= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \frac{1}{(1+t^2/n)^{(n+1)/2}}, \end{aligned}$$

correspondiente a la función de densidad de la distribución $t(n)$. Salvo la constante $\Gamma((n+1)/2)$, la última integral corresponde a la densidad $\text{gama}((n+1)/2, \lambda)$ con $\lambda = 1$. \square

El siguiente resultado es usado en estadística para efectuar estimaciones de la media de una población normal cuando la varianza es desconocida.

PROPOSICIÓN. Sea X_1, \dots, X_n una m.a. de una distribución $N(\mu, \sigma^2)$. Entonces

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Demostración. Simplemente se aplica la proposición recién demostrada a las variables aleatorias independientes

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{y} \quad \frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1).$$

\square

DISTRIBUCIÓN F. La variable aleatoria continua X tiene una distribución F de Snedecor con parámetros $n > 0$ y $m > 0$ si su función de densidad es

$$f(x) = \begin{cases} \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \left(\frac{n}{m}\right)^{n/2} x^{n/2-1} \left(1 + \frac{n}{m}x\right)^{-(n+m)/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se escribe $X \sim F(n, m)$. En la Figura 6.3 se muestra el comportamiento de esta función de densidad.

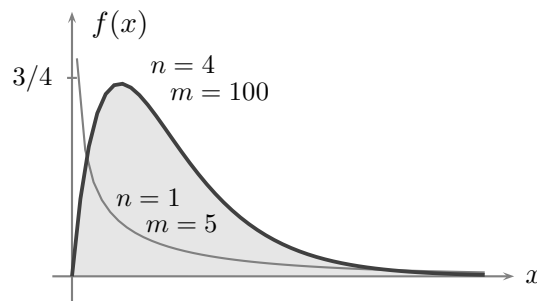


Figura 6.3: Función de densidad $F(n, m)$.

Puede demostrarse que

$$E(X) = \frac{m}{m-2}, \quad \text{para } m > 2,$$

y

$$\text{Var}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, \quad \text{para } m > 4.$$

Los siguientes dos resultados indican la forma de obtener esta distribución.

PROPOSICIÓN. Sean $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$ independientes. Entonces

$$\frac{X/n}{Y/m} \sim F(n, m).$$

Demostración. Esta afirmación se obtiene directamente de la aplicación de la fórmula para la función de densidad del cociente de dos variables aleatorias. Recuerde que para $n > 0$, $f_{X/n}(x) = n f_X(nx)$. \square

PROPOSICIÓN. Si $X \sim t(n)$, entonces $X^2 \sim F(1, n)$.

Demostración. El resultado se sigue fácilmente de la aplicación de la siguiente fórmula general. Para $x > 0$, y por la simetría de la distribución t ,

$$f_{X^2}(x) = (f_X(\sqrt{x}) + f_X(-\sqrt{x})) \frac{1}{2\sqrt{x}} = f_X(\sqrt{x}) \frac{1}{\sqrt{x}}.$$

\square

6.2. Estadísticas de orden

Dada una muestra aleatoria X_1, \dots, X_n , podemos evaluar cada una de estas variables en un punto muestral ω cualquiera y obtener una colección de números reales $X_1(\omega), \dots, X_n(\omega)$. Estos números pueden ser ordenados de menor a mayor incluyendo repeticiones. Si $X_{(i)}(\omega)$ denota el i -ésimo número ordenado, tenemos entonces la colección no decreciente de números reales

$$X_{(1)}(\omega) \leq \dots \leq X_{(n)}(\omega).$$

Ahora hacemos variar el argumento ω y lo que se obtiene son las así llamadas estadísticas de orden. Este proceso de ordenamiento resulta ser de importancia en algunas aplicaciones. Tenemos entonces la siguiente definición.

DEFINICIÓN. (ESTADÍSTICAS DE ORDEN). Sea X_1, \dots, X_n una muestra aleatoria. A las variables aleatorias ordenadas

$$\begin{aligned} X_{(1)} &= \text{mín} \{X_1, \dots, X_n\}, \\ X_{(2)} &= \text{mín} \{X_1, \dots, X_n\} \setminus \{X_{(1)}\}, \\ X_{(3)} &= \text{mín} \{X_1, \dots, X_n\} \setminus \{X_{(1)}, X_{(2)}\}, \\ &\vdots \\ X_{(n)} &= \text{máx} \{X_1, \dots, X_n\}, \end{aligned}$$

se les conoce con el nombre de estadísticas de orden. A $X_{(1)}$ se le llama primera estadística de orden, a $X_{(2)}$ se le llama segunda estadística de orden, etc. A $X_{(i)}$ se le llama i -ésima estadística de orden, $i = 1, \dots, n$.

Observe que, aunque los elementos de la muestra aleatoria son variables aleatorias independientes, las estadísticas de orden no lo son, pues deben mantener la relación $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Observe además que la i -ésima estadística de orden $X_{(i)}$ no necesariamente es igual a alguna variable de la muestra aleatoria en particular, sino que, en general, es una función de todas las variables de la muestra aleatoria.

Nuestro objetivo es encontrar algunas fórmulas relacionadas con las distribuciones de probabilidad de las estadísticas de orden cuando se conoce la distribución de las variables de la muestra aleatoria, que por simplicidad se supondrá absolutamente continua. En lo que resta del capítulo supondremos entonces que X_1, \dots, X_n es una muestra aleatoria en donde cada variable tiene función de densidad $f(x)$ y función de distribución $F(x)$.

Distribuciones individuales

Comenzamos encontrando la distribución de la primera y de la última estadística de orden de manera individual.

PROPOSICIÓN. Para $n \geq 1$,

1. $f_{X_{(1)}}(x) = nf(x) [1 - F(x)]^{n-1}$.
2. $f_{X_{(n)}}(x) = nf(x) [F(x)]^{n-1}$.

Demostración.

1. Se calcula primero la función de distribución.

$$\begin{aligned}
 F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) \\
 &= P(\min\{X_1, \dots, X_n\} \leq x) \\
 &= 1 - P(\min\{X_1, \dots, X_n\} > x) \\
 &= 1 - P(X_1 > x, \dots, X_n > x) \\
 &= 1 - [P(X_1 > x)]^n \\
 &= 1 - [1 - F(x)]^n.
 \end{aligned}$$

Entonces $f_{X_{(1)}}(x) = nf(x) [1 - F(x)]^{n-1}$.

2. Se procede de manera análoga.

$$\begin{aligned}
 F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) \\
 &= P(\max\{X_1, \dots, X_n\} \leq x) \\
 &= P(X_1 \leq x, \dots, X_n \leq x) \\
 &= [P(X_1 \leq x)]^n \\
 &= [F(x)]^n.
 \end{aligned}$$

Por lo tanto $f_{X_{(n)}}(x) = nf(x) [F(x)]^{n-1}$.

□

EJERCICIO. Compruebe que las expresiones encontradas para $f_{X_{(1)}}$ y $f_{X_{(n)}}$ son efectivamente funciones de densidad. Encuentre en particular las expresiones para estas funciones cuando las variables de la muestra tienen distribución $\text{unif}(0, 1)$. ■

Ahora se presenta el resultado general acerca de la función de densidad de la i -ésima estadística de orden.

PROPOSICIÓN. La función de densidad de la i -ésima estadística de orden es

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}.$$

Demostración. Para cada i defina la variable aleatoria

$$Y_i = 1_{(-\infty, x]}(X_i) = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{si } X_i > x, \end{cases}$$

en donde X_i es el i -ésimo elemento de la muestra aleatoria. Las variables Y_1, \dots, Y_n son independientes y cada una de ellas puede considerarse un ensayo Bernoulli con probabilidad de éxito, es decir tomar el valor 1, igual a $P(X_i \leq x) = F(x)$. Entonces la suma $Y_1 + \dots + Y_n$ corresponde al número de variables aleatorias X_i que cumplen la condición $X_i \leq x$, y por lo tanto esta suma tiene distribución $\text{bin}(n, p)$, con $p = F(x)$. Entonces

$$\begin{aligned} F_{X_{(i)}}(x) &= P(X_{(i)} \leq x) \\ &= P(Y_1 + \dots + Y_n \geq i) \\ &= \sum_{j=i}^n \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j}. \end{aligned}$$

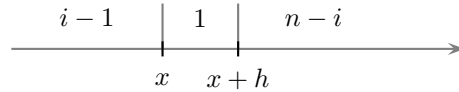
Derivando y después simplificando,

$$\begin{aligned}
 f_{X_{(i)}}(x) &= \sum_{j=i}^n \binom{n}{j} f(x)[F(x)]^{j-1}[1-F(x)]^{n-j-1} \\
 &\qquad\qquad\qquad [j(1-F(x)) - (n-j)F(x)] \\
 &= \sum_{j=i}^n \binom{n}{j} j f(x)[F(x)]^{j-1}[1-F(x)]^{n-j} \\
 &\quad - \sum_{j=i}^n \binom{n}{j} (n-j) f(x)[F(x)]^j [1-F(x)]^{n-j-1} \\
 &= \binom{n}{i} i f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}.
 \end{aligned}$$

□

EJERCICIO. Demuestre que la expresión encontrada para $f_{X_{(i)}}(x)$ es efectivamente una función de densidad. Verifique que esta densidad se reduce a las encontradas antes cuando el índice i toma los valores 1 o n . En particular, encuentre la función de densidad de la i -ésima estadística de orden suponiendo que las variables de la muestra tienen distribución $\text{unif}(0, 1)$. .

A continuación se presenta un argumento corto e intuitivo que nos lleva al mismo resultado. Sea $h > 0$ arbitrario, y considere los siguientes tres intervalos ajenos $(-\infty, x]$, $(x, x + h]$ y $(x + h, \infty)$.



La probabilidad de que $i - 1$ variables de la muestra tomen un valor en el intervalo $(-\infty, x]$, una de ellas en $(x, x + h]$, y el resto $n - i$ en $(x + h, \infty)$ es, de acuerdo a la distribución multinomial,

$$\frac{n!}{(i-1)! 1! (n-i)!} [F(x)]^{i-1} [F(x+h) - F(x)] [1 - F(x+h)]^{n-i}.$$

Esta probabilidad es aproximadamente igual a $f_{X_{(i)}}(x)h$. Dividiendo entre h , y después haciendo h tender a cero se obtiene nuevamente

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i}.$$

Sea X_1, \dots, X_n una muestra aleatoria. A la variable aleatoria $R = X_{(n)} - X_{(1)}$ se le conoce como el *rango* de la muestra. El siguiente resultado provee de una fórmula para la función de densidad de esta variable.

PROPOSICIÓN. Para $r > 0$,

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(v)f(r+v)[F(r+v) - F(v)]^{n-2} dv.$$

Demostración. Para $x < y$,

$$\begin{aligned} F_{X_{(1)}, X_{(n)}}(x, y) &= P(X_{(1)} \leq x, X_{(n)} \leq y) \\ &= P(X_{(n)} \leq y) - P(X_{(n)} \leq y, X_{(1)} > x) \\ &= [F(y)]^n - P(x < X_1 \leq y, \dots, x < X_n \leq y) \\ &= [F(y)]^n - [F(y) - F(x)]^n. \end{aligned}$$

Por lo tanto, $f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)f(x)f(y)[F(y) - F(x)]^{n-2}$, para $n \geq 2$. Ahora se usa la fórmula

$$f_{Y-X}(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u+v) dv$$

equivalente a (5.5) para la diferencia de dos variables aleatorias. Entonces para $r > 0$,

$$f_{X_{(n)}-X_{(1)}}(r) = n(n-1) \int_{-\infty}^{\infty} f(v)f(r+v)[F(r+v) - F(v)]^{n-2} dv.$$

□

EJERCICIO. Se escogen n puntos al azar con distribución uniforme en el intervalo unitario $(0, 1)$. Demuestre que la función de densidad de la distancia máxima entre cualesquiera dos puntos es

$$f(r) = \begin{cases} n(n-1)r^{n-2}(1-r) & \text{si } 0 < r < 1, \\ 0 & \text{otro caso.} \end{cases}$$

▪

Distribuciones conjuntas

Se presentan a continuación dos resultados acerca de la distribución conjunta de las estadísticas de orden. El primer resultado trata acerca de la distribución conjunta de todas ellas, después se considera la distribución conjunta de cualesquiera dos.

PROPOSICIÓN. Para $x_1 < \dots < x_n$,

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n! f(x_1) \cdots f(x_n).$$

Demostración. Se considera la función de distribución conjunta de todas las estadísticas de orden, y después se deriva n veces para encontrar la función de densidad. Para $x_1 < x_2 < \dots < x_n$,

$$F_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n).$$

Como $(X_{(2)} \leq x_2) = (x_1 < X_{(2)} \leq x_2) \cup (X_{(2)} \leq x_1)$, se obtiene la expresión

$$\begin{aligned} F_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) &= P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, X_{(n)} \leq x_n) \\ &\quad + P(X_{(1)} \leq x_1, X_{(2)} \leq x_1, \dots, X_{(n)} \leq x_n). \end{aligned}$$

Observe que el segundo sumando no depende de x_2 , así es que al tomar la derivada respecto de esta variable, este término desaparece. De manera análoga procedemos con los eventos $(X_{(3)} \leq x_3)$ hasta $(X_{(n)} \leq x_n)$. Al final se obtiene

$$\begin{aligned} f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \\ = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, x_{n-1} < X_{(n)} \leq x_n). \end{aligned}$$

Como ahora los intervalos involucrados son disjuntos, la distribución multinomial asegura que

$$\begin{aligned} P(X_{(1)} \leq x_1, x_1 < X_{(2)} \leq x_2, \dots, x_{n-1} < X_{(n)} \leq x_n) \\ = n! P(X_1 \leq x_1, x_1 < X_2 \leq x_2, \dots, x_{n-1} < X_n \leq x_n) \\ = n! F(x_1)[F(x_2) - F(x_1)] \cdots [F(x_n) - F(x_{n-1})], \end{aligned}$$

en donde la última igualdad se sigue de la independencia e idéntica distribución de las variables de la muestra. Ahora solo resta derivar para encontrar el resultado buscado, siendo más sencillo encontrar las derivadas en el orden inverso. \square

EJERCICIO. Demuestre que la expresión encontrada para la función de densidad conjunta de las estadísticas de orden es efectivamente una función de densidad multivariada. Encuentre además esta función cuando las variables de la muestra tienen distribución $\text{unif}(0, 1)$. \blacksquare

La siguiente demostración es una prueba corta pero no formal del mismo resultado. Sean $x_1 < x_2 < \cdots < x_n$, y $h > 0$ suficientemente pequeña tal que los intervalos $(x_1, x_1 + h], (x_2, x_2 + h], \dots, (x_n, x_n + h]$ son ajenos. Véase la Figura 6.4.

La probabilidad de que las variables aleatorias tomen valores, cada una de ellas, en uno y sólo uno de estos intervalos es, de acuerdo a la distribución multinomial,

$$\frac{n!}{1! \cdots 1!} [F(x_1 + h) - F(x_1)] \cdots [F(x_n + h) - F(x_n)].$$

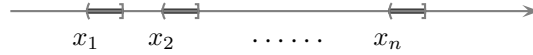


Figura 6.4:

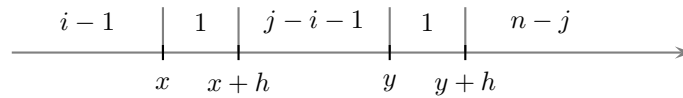


Figura 6.5:

Esta probabilidad es aproximadamente igual a $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)h^n$. Dividiendo entre h^n , y después haciendo h tender a cero se obtiene, una vez mas, $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n!f(x_1) \cdots f(x_n)$.

Ahora nos interesa encontrar una fórmula para la densidad conjunta de cualesquiera dos estadísticas de orden.

PROPOSICIÓN. Suponga $i < j$. Para $x < y$,

$$f_{X_{(i)}, X_{(j)}}(x, y) = \binom{n}{i, j-i, n-j} i(j-i) f(x)f(y) [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1 - F(y)]^{n-j}.$$

Para este resultado se presenta únicamente el argumento intuitivo usado antes. Sean $x < y$ y considere los intervalos ajenos $(-\infty, x]$, $(x, x + h]$, $(x + h, y]$, $(y, y + h]$, y $(y + h, \infty)$ para $h > 0$ suficientemente pequeña. véase la Figura 6.5.

La probabilidad de que $i - 1$ variables de la muestra tomen un valor en

$(-\infty, x]$, una de ellas en $(x, x + h]$, $j - i + 1$ variables en $(x + h, y]$, otra en $(y, y + h]$, y el resto, $n - j$ variables, tomen un valor en $(y + h, \infty)$ es, de acuerdo a la distribución multinomial,

$$\frac{n!}{(i-1)! 1! (j-i-1)! 1! (n-j)!} [F(x)]^{i-1} [F(x+h) - F(x)] [F(y) - F(x+h)]^{j-i-1} [F(y+h) - F(y)] [1 - F(y+h)]^{n-j}.$$

Esta probabilidad es aproximadamente igual a $f_{X_{(i)}, X_{(j)}}(x, y) h^2$. Dividiendo entre h^2 , y después haciendo h tender a cero se obtiene la fórmula enunciada.

EJERCICIO. Demuestre que la expresión encontrada para la función de densidad conjunta de las estadísticas de orden $X_{(i)}$ y $X_{(j)}$ es efectivamente una función de densidad bivariada. Encuentre además esta función cuando las variables de la muestra tienen distribución $\text{unif}(0, 1)$. •

Las fórmulas para las funciones de densidad de las estadísticas de orden encontradas en este capítulo se resumen en la siguiente tabla.

FÓRMULAS PARA LAS FUNCIONES DE DENSIDAD DE ALGUNAS
ESTADÍSTICAS DE ORDEN EN EL CASO ABSOLUTAMENTE CONTINUO

$$f_{X_{(1)}}(x) = n f(x) [1 - F(x)]^{n-1}$$

$$f_{X_{(n)}}(x) = n f(x) [F(x)]^{n-1}$$

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}$$

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(v) f(r+v) [F(r+v) - F(v)]^{n-2} dv, \\ \text{para } r > 0 \text{ en donde } R = X_{(n)} - X_{(1)}$$

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = n! f(x_1) \cdots f(x_n), \text{ para } x_1 < \cdots < x_n$$

$$f_{X_{(i)}, X_{(j)}}(x, y) = \binom{n}{i, j-i, n-j} i(j-i) f(x) f(y) [F(x)]^{i-1} [F(y) - F(x)]^{j-i-1} [1 - F(y)]^{n-j}, \\ \text{para } x < y \text{ e } i < j$$

6.3. Ejercicios

Media y varianza muestral

472. Sea X_1, \dots, X_n una muestra aleatoria de una distribución con media μ y varianza σ^2 . Demuestre que $E(\bar{X}) = \mu$ y $E(S^2) = \sigma^2$. Estos resultados son de utilidad en estadística y muestran que \bar{X} y S^2 son *estimadores insesgados* para la media y varianza de la distribución.
473. Sea X_1, \dots, X_n una m.a. de una distribución con media μ y varianza σ^2 . Demuestre que $\text{Var}(\bar{X}) = \sigma^2/n$. ¿Cuánto vale $\text{Var}(S^2)$?
474. Sea X_1, \dots, X_n una m.a. de una distribución $\text{Ber}(p)$. Demuestre que las estadísticas \bar{X} y S^2 no son independientes.

Distribución χ^2

475. Demuestre que la función de densidad de la distribución $\chi^2(n)$ efectivamente lo es. En particular, compruebe que la distribución $\chi^2(n)$, con $n = 2$, se reduce a la distribución $\exp(\lambda)$ con $\lambda = 1/2$.
476. Demuestre que la distribución $\text{gama}(n/2, \lambda)$, con $\lambda = 1/2$, se reduce a la distribución $\chi^2(n)$.
477. Sea X con distribución $\chi^2(n)$. Demuestre que
- $E(X) = n$.
 - $E(X^m) = 2^m \Gamma(m + n/2)/\Gamma(n/2)$, para $m = 1, 2, \dots$
 - $\text{Var}(X) = 2n$.
478. Sean X_1, \dots, X_n independientes cada una con distribución $N(\mu, \sigma^2)$. Demuestre que

$$\frac{(\bar{X} - \mu)^2}{\sigma^2/n} \sim \chi^2(1).$$

479. Sean X_1, \dots, X_n independientes cada una con distribución normal estándar. Demuestre que

$$\sum_{i=1}^n X_i^2 \sim \chi^2(n).$$

480. Sean X_1, \dots, X_n independientes tales que cada variable X_i tiene distribución $N(\mu_i, \sigma_i^2)$ para $i = 1, \dots, n$. Demuestre que

$$\sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \sim \chi^2(n).$$

481. Sean X y Y independientes ambas con distribución normal estándar. Sean $R = \sqrt{X^2 + Y^2}$ y $\theta = \tan^{-1}(Y/X)$. Demuestre que

- R^2 tiene distribución $\chi^2(n)$ con $n = 2$ grados de libertad.
- $\tan \theta$ tiene distribución Cauchy.
- R y θ son independientes.

Distribución t

482. Demuestre que la función de densidad de una variable aleatoria X con distribución $t(n)$ efectivamente lo es. Demuestre además que esta función tiene un máximo en $x = 0$ y que

- $E(X) = 0$.
- $\text{Var}(X) = n/(n-2)$, para $n > 2$.

Compruebe además que esta distribución se reduce a la distribución Cauchy cuando el valor del parámetro n es uno.

483. Demuestre que la distribución $t(n+1)$ tiene momentos finitos de orden menor o igual a n , pero ningún otro momento de orden superior.

Distribución F

484. Demuestre que la función de densidad de una variable aleatoria X con distribución $F(n, m)$ efectivamente lo es. Demuestre además que
- a) $E(X) = m/(m - 2)$, para $m > 2$.
 - b) $\text{Var}(X) = \frac{2m^2(m + n - 2)}{n(m - 2)^2(m - 4)}$, para $m > 4$.
485. Sea X con distribución $F(n, m)$. Demuestre que $Y = 1/X$ tiene distribución $F(m, n)$, observe el cambio en el orden de los parámetros. Este resultado es útil para obtener valores de F que no aparecen en tablas de esta distribución que son comunes en textos de estadística.
486. Sea X con distribución $F(n, m)$. Demuestre que cuando m tiende a infinito la función de densidad de nX converge a la función de densidad de la distribución $\chi^2(n)$.

Estadísticas de orden: distribuciones individuales

487. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Demuestre que la i -ésima estadística de orden tiene distribución $\text{beta}(i, n + 1 - i)$. Encuentre por lo tanto su esperanza y varianza.
488. Sea X_1, \dots, X_n una m.a. de una distribución $\exp(\lambda)$. Encuentre la función de densidad de la i -ésima estadística de orden.
489. Sean $X_{(1)}, X_{(2)}$ las estadísticas de orden de una m.a. de tamaño dos de una distribución $N(\mu, \sigma^2)$. Demuestre que $E[X_{(1)}] = \mu - \sigma/\sqrt{\pi}$ y calcule $E[X_{(2)}]$.
490. Sea X_1, \dots, X_n una m.a. de una distribución $F(x)$. Sea x un número real cualquiera, y para cada $i = 1, \dots, n$ defina $Y_i = 1_{(-\infty, x]}(X_i)$. Demuestre que las variables Y_1, \dots, Y_n son independientes, y cada una

de ellas tiene distribución $\text{Ber}(p)$, con $p = F(x)$. Este hecho fue utilizado en el procedimiento para encontrar la función de densidad de la i -ésima estadística de orden.

491. Sean X y Y absolutamente continuas e independientes. Defina $V = \max\{X, Y\}$. Demuestre que
- $F_V(v) = F_X(v)F_Y(v)$.
 - $f_V(v) = F_X(v)f_Y(v) + f_X(v)F_Y(v)$.
 - $f_V(v) = 2F(v)f(v)$, cuando X y Y tienen la misma distribución.
492. Use el ejercicio anterior para encontrar la función de densidad del máximo de dos variables aleatorias independientes cada una de ellas con distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.
493. Sean X y Y absolutamente continuas e independientes. Defina $U = \min\{X, Y\}$. Demuestre que
- $F_U(u) = 1 - [1 - F_X(u)][1 - F_Y(u)]$.
 - $f_U(u) = [1 - F_X(u)]f_Y(u) + f_X(u)[1 - F_Y(u)]$.
 - $f_U(u) = 2[1 - F(u)]f(u)$, cuando X y Y tienen la misma distribución.
494. Use el ejercicio anterior para encontrar la función de densidad del mínimo de dos variables aleatorias independientes cada una de ellas con distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.
495. Sean X_1, \dots, X_n variables aleatorias independientes en donde X_k tiene distribución $\exp(\lambda_k)$, para $k = 1, \dots, n$. Demuestre que la variable $\min\{X_1, \dots, X_n\}$ tiene distribución $\exp(\lambda_1 + \dots + \lambda_n)$, y que $P(X_k = \min\{X_1, \dots, X_n\}) = \lambda_k / (\lambda_1 + \dots + \lambda_n)$.

Estadísticas de orden: distribuciones conjuntas

496. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(1)}$, encontrando nuevamente que

$$f_{X_{(1)}}(x) = nf(x)[1 - F(x)]^{n-1}.$$

497. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(n)}$, encontrando nuevamente que

$$f_{X_{(n)}}(x) = nf(x)[F(x)]^{n-1}.$$

498. A partir de la fórmula para $f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n)$, calcule la función de densidad marginal de $X_{(i)}$, para $i = 1, \dots, n$, encontrando nuevamente que

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i}.$$

499. A partir de la fórmula para $f_{X_{(i)}, X_{(j)}}(x, y)$, calcule la función de densidad marginal de $X_{(i)}$, encontrando nuevamente que

$$f_{X_{(i)}}(x) = \binom{n}{i} i f(x)[F(x)]^{i-1}[1 - F(x)]^{n-i}.$$

500. Sea X_1, \dots, X_n una m.a. de una distribución unif(-1, 1). Encuentre la función de densidad de

- a) $X_{(1)}$ y $X_{(2)}$ conjuntamente.
- b) $R = X_{(n)} - X_{(1)}$.

501. MEDIANA MUESTRAL. La mediana de una muestra aleatoria X_1, \dots, X_n , denotada por $\text{Med}(X_1, \dots, X_n)$, se define del siguiente modo. Considere las estadísticas de orden $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, entonces

$$\text{Med}(X_1, \dots, X_n) = \begin{cases} X_{(\frac{n+1}{2})} & \text{si } n \text{ es impar,} \\ \frac{1}{2} [X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & \text{si } n \text{ es par.} \end{cases}$$

Encuentre la función de densidad de la mediana de una muestra aleatoria de la distribución $\text{unif}(0, 1)$, primero suponiendo que el tamaño de la muestra n es impar, y después para n par.

502. Sea X_1, \dots, X_n una m.a. de una distribución $\text{unif}(0, 1)$. Calcule el coeficiente de correlación entre $X_{(i)}$ y $X_{(j)}$.
503. Sea X_1, \dots, X_n una m.a. de una distribución continua $F(x)$ con función de densidad $f(x)$. Demuestre directamente que para $x < y$,

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)f(x)f(y)[F(y) - F(x)]^{n-2}.$$

504. Encuentre la función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$ para una m.a. de tamaño n de una distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.
505. Calcule la covarianza entre $X_{(1)}$ y $X_{(n)}$ para una m.a. de tamaño n de una distribución: a) $\text{unif}(0, 1)$. b) $\exp(\lambda)$.

CAPÍTULO 7

Convergencia

En este capítulo se presenta una introducción al tema de convergencia de variables aleatorias. Históricamente este tema surge a través de ciertas preguntas que se formularon acerca del comportamiento del promedio de variables aleatorias $\frac{1}{n} \sum_{i=1}^n X_i$ cuando n crece a infinito. En la última parte del texto estudiaremos algunos resultados importantes sobre este comportamiento limite particular. En este capítulo estudiaremos distintas formas en que una sucesión infinita de variables aleatorias puede converger de manera general. En la mayoría de las situaciones que consideraremos supondremos que existe un espacio de probabilidad (Ω, \mathcal{F}, P) en donde una sucesión infinita de variables aleatorias X_1, X_2, \dots están todas ellas definidas.

7.1. Tipos de convergencia

Convergencia puntual

Sea X_1, X_2, \dots una sucesión infinita de variables aleatorias. Al evaluar cada una de estas variables en un elemento ω se obtiene la sucesión numérica $X_1(\omega), X_2(\omega), \dots$. Suponga que esta sucesión converge a un cierto número real denotado por $X(\omega)$. Si lo anterior se cumple para todos y cada uno

de los elementos de Ω , entonces se dice que la sucesión de variables aleatorias converge puntualmente, y su límite es la función $X : \Omega \rightarrow \mathbb{R}$ definida naturalmente por $X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)$. Se ha demostrado antes que en esta situación la función límite X es efectivamente una variable aleatoria. Formalmente se tiene entonces la siguiente definición.

DEFINICIÓN. (CONVERGENCIA PUNTUAL). La sucesión de variables aleatorias X_1, X_2, \dots converge puntualmente a X si para cada ω en Ω ,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

EJEMPLO. Considere el espacio medible $([0, 1], \mathcal{B}[0, 1])$, y defina la sucesión de variables aleatorias continuas $X_n(\omega) = \omega^n$. Como en este caso el espacio muestral es un subconjunto de números reales, podemos graficar las variables aleatorias como en la Figura 7.1.

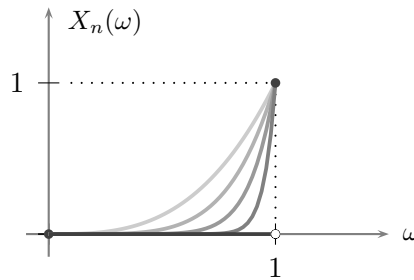


Figura 7.1: Gráfica de la variable aleatoria $X_n(\omega) = \omega^n$.

Entonces para cada $\omega \in [0, 1)$, la sucesión numérica $X_n(\omega)$ converge a 0, mientras que para $\omega = 1$, y para cualquier valor de n , $X_n(\omega) = 1$. De esta manera la sucesión converge puntualmente a la variable aleatoria

$$X(\omega) = \begin{cases} 0 & \text{si } \omega \in [0, 1), \\ 1 & \text{si } \omega = 1. \end{cases}$$

EJERCICIO. Considere el espacio medible $(\mathbb{N}, 2^{\mathbb{N}})$. Determine si existe convergencia puntual para cada una de las siguientes sucesiones de variables aleatorias discretas. En caso afirmativo encuentre la variable aleatoria límite. a) $X_n(\omega) = \omega \bmod n$ b) $X_n(\omega) = \min\{\omega, n\}$ c) $X_n(\omega) = \max\{\omega, n\}$

Una sucesión de variables aleatorias es entonces una sucesión de funciones, pero a diferencia de la situación que se estudia en los cursos de análisis matemático, el dominio de definición de estas funciones, es decir, el espacio muestral en este caso, no tiene una estructura algebraica excepto la dada por la σ -álgebra y la medida de probabilidad. La forma en la que se utiliza esta medida de probabilidad es la que determina los distintos tipos de convergencia.

Convergencia casi segura

En algunas situaciones la convergencia puntual resulta ser una condición muy fuerte pues se pide la convergencia de la sucesión evaluada en todos y cada uno de los elementos del espacio muestral. Se puede ser menos estricto y pedir, por ejemplo, que la convergencia se verifique en todo el espacio Ω excepto en un subconjunto de probabilidad cero.

DEFINICIÓN. (CONVERGENCIA CASI SEGURA). La sucesión de variables aleatorias X_1, X_2, \dots converge casi seguramente a la variable X , si

$$P\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1.$$

Es decir, en la convergencia casi segura se permite que para algunos valores de ω , la sucesión numérica $X_1(\omega), X_2(\omega), \dots$ pueda no converger, sin embargo el subconjunto de Ω en donde esto suceda debe tener probabilidad

cero. Para indicar la convergencia casi segura se escribe $X_n \xrightarrow{c.s.} X$, o bien $\lim_{n \rightarrow \infty} X_n = X$ *c.s.* A menudo se utiliza el término convergencia *casi dondequiera*, o bien convergencia *casi siempre* para denotar este tipo de convergencia. Observe que omitiendo el argumento ω , la condición para la convergencia casi segura se escribe en la forma más corta: $P(\lim_{n \rightarrow \infty} X_n = X) = 1$, o simplemente $P(X_n \rightarrow X) = 1$. Es posible demostrar que el conjunto $\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}$ es medible de modo que tiene sentido aplicar la probabilidad, al respecto véase el ejercicio 506. Puede también demostrarse que bajo este tipo de convergencia, el límite es único casi seguramente, es decir, si X_n converge a X *c.s.* y también converge a Y *c.s.*, entonces $X = Y$ casi seguramente.

EJEMPLO. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$ con P la medida uniforme, es decir, la medida de probabilidad de un intervalo es su longitud. Defina la sucesión de variables aleatorias como se muestran en la Figura 7.2.

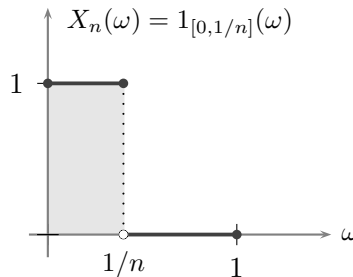


Figura 7.2: Gráfica de la variable aleatoria $X_n(\omega) = 1_{[0, 1/n]}(\omega)$.

Es decir, la variable X_n tiene distribución Bernoulli con parámetro $p = 1/n$, y converge casi seguramente a la variable aleatoria constante cero. Para demostrar esto se necesita verificar que $P(X_n \rightarrow 0) = 1$. Pero esta igualdad es evidente a partir del hecho de que el conjunto $\{\omega \in \Omega : X_n(\omega) \rightarrow 0\}$ es el intervalo $(0, 1]$, el cual tiene probabilidad uno. El punto $\omega = 0$ es el único punto muestral para el cual $X_n(\omega)$ no converge a cero. Esto demuestra que

$$X_n \xrightarrow{c.s.} 0. \quad \cdot$$

EJERCICIO. Sea A un evento cualquiera. Demuestre que la siguiente sucesión de variables aleatorias no converge para ningún ω en Ω .

$$X_n = \begin{cases} 1_A & \text{si } n \text{ es par,} \\ 1_{A^c} & \text{si } n \text{ es impar.} \end{cases} \quad \cdot$$

Convergencia en probabilidad

Un tipo de convergencia aún menos restrictiva que la convergencia casi segura es la convergencia en probabilidad la cual se define a continuación.

DEFINICIÓN. (CONVERGENCIA EN PROBABILIDAD). La sucesión de variables aleatorias X_1, X_2, \dots converge en probabilidad a X , si para cada $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\} = 0.$$

Para denotar la convergencia en probabilidad se escribe $X_n \xrightarrow{p} X$, y omitiendo el argumento ω la condición se escribe $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$. Nuevamente puede comprobarse que el límite es único casi seguramente.

EJEMPLO. Considere el espacio de probabilidad $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida uniforme. Defina la sucesión de eventos

$$\begin{aligned} A_1 &= (0, 1/2), & A_2 &= (1/2, 1), \\ A_3 &= (0, 1/3), & A_4 &= (1/3, 2/3), & A_5 &= (2/3, 1), \\ A_6 &= (0, 1/4), & A_7 &= (1/4, 2/4), & A_8 &= (2/4, 3/4), & A_9 &= (3/4, 1), \\ &\dots\dots \end{aligned}$$

Sea $X_n = 1_{A_n}$. Las gráficas de estas primeras variables aleatorias se muestran en la Figura 7.3. Entonces $X_n \xrightarrow{p} 0$ pues para cualquier $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|X_n - 0| > \epsilon) = \lim_{n \rightarrow \infty} P(A_n) = 0.$$

Por otro lado observe que esta sucesión de variables aleatorias no converge casi seguramente pues el conjunto $\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) \text{ existe}\}$ es vacío.

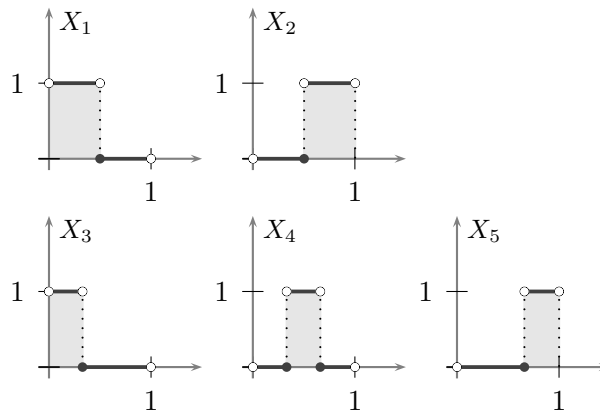


Figura 7.3: Gráficas de las primeras variables aleatorias $X_n = 1_{A_n}$.

En algunos casos la aplicación de la desigualdad de Chebyshev resulta útil para demostrar este tipo de convergencia como se muestra a continuación.

EJERCICIO. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes cada una de ellas con distribución $N(\mu, \sigma^2)$ y defina el promedio $S_n = \frac{1}{n} \sum_{i=1}^n X_i$. Use la desigualdad de Chebyshev para demostrar que $S_n \xrightarrow{p} \mu$. Observe que el mismo argumento funciona para cualquier sucesión de variables aleatorias independientes idénticamente distribuidas con varianza finita.

Convergencia en media

En este tipo de convergencia se usa la esperanza para determinar la cercanía entre dos variables aleatorias.

DEFINICIÓN. (CONVERGENCIA EN MEDIA). La sucesión de variables aleatorias integrables X_1, X_2, \dots converge en media a la variable aleatoria integrable X si

$$\lim_{n \rightarrow \infty} E|X_n - X| = 0.$$

A este tipo de convergencia también se le llama *convergencia en L^1* y se le denota por $X_n \xrightarrow{m} X$, o $X_n \xrightarrow{L^1} X$.

A partir de la definición de convergencia en media es inmediato preguntarse si de allí se sigue la convergencia de la sucesión de medias. La respuesta es afirmativa.

EJERCICIO. Use la desigualdad de Jensen para demostrar que si $X_n \xrightarrow{m} X$, entonces $E(X_n) \rightarrow E(X)$. ▪

Convergencia en media cuadrática

Nuevamente usando el concepto de esperanza pero ahora aplicado al segundo momento se tiene la convergencia en media cuadrática. Más adelante demostraremos que la convergencia en media cuadrática implica la convergencia en media.

DEFINICIÓN. (CONVERGENCIA EN MEDIA CUADRÁTICA). La sucesión de variables aleatorias X_1, X_2, \dots converge en media cuadrática a X , si

$$\lim_{n \rightarrow \infty} E|X_n - X|^2 = 0.$$

En este tipo de convergencia se presupone que tanto los elementos de la sucesión como el límite mismo son variables aleatorias con segundo momento finito. A este tipo de convergencia también se le llama *convergencia en L^2* , y se le denota por $X_n \xrightarrow{m.c.} X$, o $X_n \xrightarrow{L^2} X$.

En general puede definirse la convergencia en L^k , para cada entero $k \geq 1$, cuando se cumple la condición $E|X_n - X|^k \rightarrow 0$. Resulta que mientras mayor es el valor de k , más restrictiva es la condición de convergencia.

Convergencia en distribución

Este es el tipo de convergencia menos restrictiva de todas las mencionadas. En contextos más generales se le llama también *convergencia débil*.

DEFINICIÓN. (CONVERGENCIA EN DISTRIBUCIÓN). La sucesión de variables aleatorias X_1, X_2, \dots converge en distribución a X , si para todo punto x en donde la función $F_X(x)$ es continua, se cumple que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

En este caso se escribe $X_n \xrightarrow{d} X$, o $F_{X_n} \xrightarrow{d} F_X$, o bien $X_n \xrightarrow{d} F_X$. Por ejemplo, si la distribución límite es la distribución normal estándar, puede escribirse $X_n \xrightarrow{d} N(0, 1)$. Observe que para este tipo de convergencia se hace

uso sólomente de las funciones de distribución y por lo tanto las variables aleatorias correspondientes pueden estar definidas en distintos espacios de probabilidad. La unicidad del límite no se da en el sentido casi seguro como en los anteriores tipos de convergencia, sino en el sentido más débil de igualdad de distribuciones.

EJEMPLO. Considere la sucesión X_1, X_2, \dots , en donde cada X_n tiene distribución $N(0, \sigma^2/n)$. Demostraremos que $X_n \xrightarrow{d} 0$. Como

$$F_{X_n}(x) = \frac{1}{\sqrt{2\pi\sigma^2/n}} \int_{-\infty}^x e^{-u^2/2(\sigma^2/n)} du,$$

e interpretando esta integral como el área bajo la curva de una función de densidad normal con media cero y varianza σ^2/n , puede comprobarse que

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1/2 & \text{si } x = 0, \\ 1 & \text{si } x > 0. \end{cases}$$

Gráficamente la distribución límite se muestra en la Figura 7.4. Observe que la variable aleatoria constante $X = 0$ tiene función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 & \text{si } x \geq 0. \end{cases}$$

Tenemos entonces que $X_n \xrightarrow{d} 0$, pues $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ para todo punto x donde $F_X(x)$ es continua, esto es, para todo x en el conjunto $\mathbb{R} \setminus \{0\}$. Observe que las funciones $F_{X_n}(x)$ no convergen a $F(x)$ cuando $x = 0$.

▪

En la siguiente sección demostraremos que la convergencia en probabilidad implica la convergencia en distribución. El recíproco en general es falso excepto cuando el límite es una constante. Este es el contenido del siguiente resultado el cual será usado más adelante para demostrar la ley débil de los grandes números.

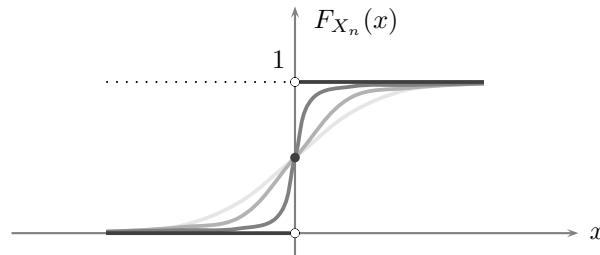


Figura 7.4: Sucesión y límite de las funciones de distribución $F_{X_n}(x)$.

PROPOSICIÓN. Sea c una constante. Si $X_n \xrightarrow{d} c$, entonces $X_n \xrightarrow{p} c$.

Demostración. La función de distribución de la variable aleatoria constante c es

$$F(x) = \begin{cases} 0 & \text{si } x < c, \\ 1 & \text{si } x \geq c, \end{cases}$$

que tiene un único punto de discontinuidad en $x = c$. Suponga entonces que $F_{X_n}(x) \rightarrow F(x)$ para $x \neq c$. Para cualquier $\epsilon > 0$ se tiene que

$$\begin{aligned} P(|X_n - c| \geq \epsilon) &= P(X_n \leq c - \epsilon) + P(X_n \geq c + \epsilon) \\ &\leq P(X_n \leq c - \epsilon) + P(X_n > c + \epsilon/2) \\ &= F_{X_n}(c - \epsilon) + 1 - F_{X_n}(c + \epsilon/2). \end{aligned}$$

De modo que $\lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) = F(c - \epsilon) + 1 - F(c + \epsilon/2) = 0$. \square

A manera de resumen y sin mayores precisiones, se presenta en la siguiente tabla las definiciones de los distintos tipos de convergencia mencionados. En la siguiente sección se estudian las relaciones entre estos tipos de convergencia.

CONVERGENCIA	DEFINICIÓN
puntual	$X_n(\omega) \rightarrow X(\omega)$ para cada ω en Ω .
casi segura	$P(X_n \rightarrow X) = 1$.
en media	$E X_n - X \rightarrow 0$.
en media cuadrática	$E X_n - X ^2 \rightarrow 0$.
en probabilidad	$P(X_n - X > \epsilon) \rightarrow 0$.
en distribución	$F_{X_n}(x) \rightarrow F_X(x)$ en puntos de continuidad x de F_X .

7.2. Relaciones entre los tipos de convergencia

En esta sección se establecen algunas relaciones generales entre los tipos de convergencia de variables aleatorias mencionados en la sección anterior. En la Figura 7.5 se ilustran de manera gráfica estas relaciones.

En este diagrama la contención se interpreta como implicación, por ejemplo, la convergencia casi segura implica la convergencia en probabilidad, y ésta a su vez implica la convergencia en distribución. Estos y otros resultados se demuestran a continuación.

PROPOSICIÓN. Convergencia c.s. \Rightarrow convergencia en prob.

Demostración. Sea $\epsilon > 0$. Para cada natural n defina los eventos

$$A_n = \bigcup_{k=n}^{\infty} (|X_k - X| > \epsilon).$$

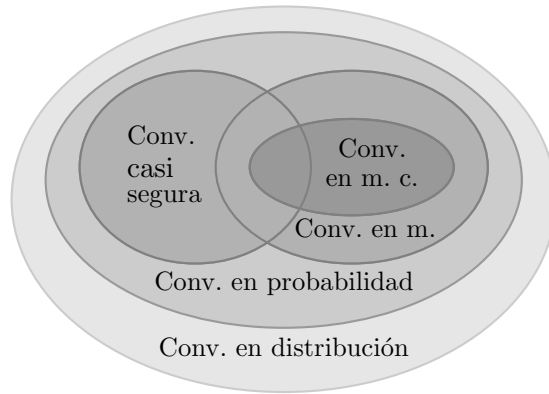


Figura 7.5: Relación entre los tipos de convergencia.

Esta sucesión es decreciente y su límite es entonces la intersección de todos los eventos. Como $(|X_n - X| > \epsilon) \subseteq A_n$, entonces $P(|X_n - X| > \epsilon) \leq P(A_n)$. Por lo tanto,

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) &\leq \lim_{n \rightarrow \infty} P(A_n) \\
 &= P\left(\lim_{n \rightarrow \infty} A_n\right) \\
 &= P\left(\bigcap_{n=1}^{\infty} A_n\right) \\
 &= P(|X_n - X| > \epsilon, \text{ para cada } n \geq 1) \\
 &= P\left(\lim_{n \rightarrow \infty} X_n \neq X\right) \\
 &= 0.
 \end{aligned}$$

□

El recíproco de la proposición anterior es, en general, falso, es decir, la convergencia en probabilidad no implica necesariamente la convergencia casi

siempre. Para comprobar esta afirmación se proporciona a continuación un ejemplo.

EJEMPLO. (EN GENERAL, CONV. EN PROB. $\not\Rightarrow$ CONV. C.S.). Considere el espacio de probabilidad $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida uniforme. Defina nuevamente la sucesión de eventos $A_1 = (0, 1/2)$, $A_2 = (1/2, 1)$, $A_3 = (0, 1/3)$, $A_4 = (1/3, 2/3)$, $A_5 = (2/3, 1)$, $A_6 = (0, 1/4)$, $A_7 = (1/4, 2/4)$, $A_8 = (2/4, 3/4)$, $A_9 = (3/4, 1), \dots$ y con ellos las variables aleatorias $X_n = 1_{A_n}$, cuyas gráficas aparecen en la Figura 7.3. Hemos comprobado antes que $X_n \xrightarrow{p} 0$, sin embargo la sucesión no converge casi seguramente pues $X_n(\omega)$ no converge para ningún ω . ■

EJEMPLO. (EN GENERAL, CONV. EN MEDIA $\not\Rightarrow$ CONVERGENCIA C.S.). Considere la sucesión de variables X_n del ejemplo anterior. Entonces $X_n \xrightarrow{m} 0$ pues $E|X_n - 0| = P(A_n) \rightarrow 0$. Sin embargo esta sucesión no converge c.s. ■

El ejemplo anterior sirve también para mostrar que, en general, la convergencia en media cuadrática no implica la convergencia casi segura. En este ejemplo se cumple que $E|X_n - 0|^2 \rightarrow 0$, y sin embargo X_n no converge a 0 c.s.

EJEMPLO. (EN GENERAL, CONV. C.S. $\not\Rightarrow$ CONV. EN MEDIA). Considere el espacio $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida de probabilidad uniforme. Defina la sucesión $X_n = n 1_{(0, 1/n)}$. Entonces X_n converge a cero casi seguramente pues $P(\lim X_n = 0) = P(\Omega) = 1$. Sin embargo no hay convergencia en media pues $E|X_n - 0| = E(X_n) = 1 \not\rightarrow 0$. ■

Este ejemplo puede ser usado también para demostrar que la convergencia casi segura no implica necesariamente la convergencia en media cuadrática.

PROPOSICIÓN. Convergencia en m.c. \Rightarrow convergencia en media.

Demostración. La desigualdad de Jensen establece que para u convexa,

$$u(E(X)) \leq E(u(X)).$$

Tomando $u(x) = x^2$ se obtiene $E^2|X_n - X| \leq E|X_n - X|^2$, de donde se sigue el resultado. Alternativamente la última desigualdad es consecuencia de la desigualdad de Cauchy-Schwarz. \square

EJEMPLO. (EN GENERAL, CONV. EN MEDIA $\not\Rightarrow$ CONV. EN M.C.) Sea $X_n = n 1_{(0,1/n^2)}$ sobre el espacio $((0,1), \mathcal{B}(0,1), P)$, con P la medida uniforme. Entonces X_n converge a cero en media pues $E|X_n - 0| = E(X_n) = 1/n \rightarrow 0$. Sin embargo, no hay convergencia en media cuadrática pues $E|X_n - 0|^2 = E(X_n^2) = 1 \not\rightarrow 0$. \cdot

PROPOSICIÓN. Convergencia en media \Rightarrow convergencia en prob.

Demostración. Para cada $\epsilon > 0$ defina el evento $A_n = (|X_n - X| > \epsilon)$. Entonces

$$\begin{aligned} E|X_n - X| &= E(|X_n - X| 1_{A_n}) + E(|X_n - X| 1_{A_n^c}) \\ &\geq E(|X_n - X| 1_{A_n}) \\ &\geq \epsilon P(|X_n - X| > \epsilon). \end{aligned}$$

Por hipótesis, el lado izquierdo tiende a cero cuando n tiende a infinito. Por lo tanto $P(|X_n - X| > \epsilon) \rightarrow 0$. \square

El recíproco del resultado anterior es, en general, falso.

EJEMPLO. (EN GENERAL, CONV. EN PROB. $\not\Rightarrow$ CONV. EN MEDIA). Considere nuevamente el espacio $((0,1), \mathcal{B}(0,1), P)$, con P la medida uniforme, y defina las variables $X_n = n 1_{(0,1/n)}$. Entonces X_n converge en probabilidad a cero pues para cualquier $\epsilon > 0$, $P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = 1/n \rightarrow 0$.

Sin embargo, la sucesión no converge en media pues $E|X_n - 0| = E(X_n) = 1 \not\rightarrow 0$. \blacksquare

PROPOSICIÓN. Convergencia en prob. \Rightarrow convergencia en dist.

Demostración. Suponga que $X_n \xrightarrow{p} X$, y sea x un punto de continuidad de $F_X(x)$. Para cualquier $\epsilon > 0$,

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, |X_n - X| \leq \epsilon) + P(X_n \leq x, |X_n - X| > \epsilon) \\ &\leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon). \end{aligned}$$

Por hipótesis el segundo sumando del lado derecho tiende a cero cuando n tiende a infinito. Entonces para cualquier $\epsilon > 0$,

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \epsilon).$$

Por la continuidad lateral,

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

Ahora se demuestra la desigualdad inversa. Para cualquier $\epsilon > 0$

$$\begin{aligned} F_X(x - \epsilon) &= P(X \leq x - \epsilon) \\ &= P(X \leq x - \epsilon, |X_n - X| \leq \epsilon) + P(X \leq x - \epsilon, |X_n - X| > \epsilon) \\ &\leq P(X_n \leq x) + P(|X_n - X| > \epsilon). \end{aligned}$$

Nuevamente el segundo sumando tiende a cero cuando n tiende a infinito. Entonces

$$F_X(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

Por la continuidad en x ,

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x).$$

En resumen,

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x).$$

□

El recíproco de la proposición anterior no siempre es válido, es decir, la convergencia en distribución no siempre implica la convergencia en probabilidad.

EJEMPLO. (EN GENERAL, CONV. EN DIST. $\not\Rightarrow$ CONV. EN PROB.) Sea X con distribución normal estándar, y sea

$$X_n = \begin{cases} X & \text{si } n \text{ es par,} \\ -X & \text{si } n \text{ es impar.} \end{cases}$$

Entonces claramente cada una de las variable X_n también tiene distribución normal estándar y por lo tanto para cualquier número real x , $F_{X_n}(x) \rightarrow F_X(x)$, es decir, $X_n \xrightarrow{d} X$. Sin embargo la sucesión no converge en probabilidad a X , pues para valores impares de n y para valores pequeños de $\epsilon > 0$, $P(|X_n - X| > \epsilon) = P(2|X| > \epsilon) > 1/2$. Lo anterior demuestra que $\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) \neq 0$. ■

Esto concluye la verificación y ejemplos de todas las implicaciones y no implicaciones que se derivan del diagrama de la Figura 7.5. El lector interesado en profundizar los temas aquí expuestos puede consultar el capítulo 5 del libro de Karr [18], o el excelente texto de Gut [13], así como los textos clásicos de teoría de la medida [5] o [14], por ejemplo. Los resultados de convergencia en espacios de probabilidad aquí mencionados pueden no ser válidos en espacios de medida más generales.

7.3. Dos resultados importantes de convergencia

Sea X_1, X_2, \dots una sucesión de variables aleatorias con esperanza finita. Suponga que X_n converge casi seguramente a X . Es natural preguntarse si la sucesión de números $E(X_n)$ converge a $E(X)$. Tal convergencia numérica equivaldría a poder intercambiar las operaciones de límite y esperanza, es decir,

$$\lim_{n \rightarrow \infty} E(X_n) = E(\lim_{n \rightarrow \infty} X_n).$$

Por ejemplo, considere el espacio $((0, 1), \mathcal{B}(0, 1), P)$, con P la medida de probabilidad uniforme. Hemos considerado antes la sucesión de variables aleatorias $X_n = n 1_{(0, 1/n)}$, cuyo límite es $X = 0$ casi seguramente. Sin embargo $E(X_n)$ es siempre 1 y no converge a $E(X) = 0$. Este es un ejemplo sencillo en donde no es válido intercambiar la esperanza y el límite. En esta sección se estudian dos resultados que establecen condiciones bajo las cuales es válido este intercambio.

TEOREMA DE CONVERGENCIA MONÓTONA. Sea $0 \leq X_1 \leq X_2 \leq \dots$ una sucesión de variables aleatorias convergente casi seguramente a una variable X . Entonces

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Demostración. Como $0 \leq X_n \leq X$, entonces $0 \leq E(X_n) \leq E(X)$. Por lo tanto

$$\lim_{n \rightarrow \infty} E(X_n) \leq E(X).$$

Ahora resta demostrar la desigualdad contraria. Primero se aproxima a X de la siguiente forma. Sea $\epsilon > 0$ arbitrario, y para cada entero $k \geq 0$ defina el evento $A_k = (k\epsilon \leq X < (k+1)\epsilon)$. Esta es una colección de eventos disjuntos dos a dos, cuya unión es Ω . Defina ahora la variable aleatoria

discreta aproximante

$$Y(\omega) = k\epsilon \quad \text{si} \quad k\epsilon \leq X(\omega) < (k+1)\epsilon.$$

Observe que Y aproxima a X de la forma: $Y \leq X < Y + \epsilon$. O bien $X - \epsilon < Y \leq X$. Por lo tanto, $E(X) - \epsilon \leq E(Y) \leq E(X)$. Para cada número natural n defina el evento $B_n = (X_n \geq Y)$. No es difícil comprobar que $B_n \nearrow \Omega$. Por lo tanto, para k fijo, $A_k \cap B_n \nearrow A_k$ cuando $n \rightarrow \infty$, y entonces $P(A_k \cap B_n) \nearrow P(A_k)$. Ahora considere la variable aleatoria discreta $Y 1_{B_n}$ dada por

$$Y 1_{B_n}(\omega) = \begin{cases} Y(\omega) & \text{si } \omega \in B_n, \\ 0 & \text{si } \omega \notin B_n. \end{cases}$$

Entonces $0 \leq Y 1_{B_n} \leq X_n$, y por lo tanto $0 \leq E(Y 1_{B_n}) \leq E(X_n)$. Entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} E(X_n) &\geq \lim_{n \rightarrow \infty} E(Y 1_{B_n}) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} E(Y 1_{B_n \cap A_k}) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} k\epsilon P(B_n \cap A_k) \\ &\geq \lim_{n \rightarrow \infty} \sum_{k=0}^m k\epsilon P(B_n \cap A_k) \\ &= \sum_{k=0}^m k\epsilon P(A_k). \end{aligned}$$

Como esta desigualdad es válida para cualquier $m \geq 0$, se obtiene

$$\lim_{n \rightarrow \infty} E(X_n) \geq \sum_{k=0}^{\infty} k\epsilon P(A_k) = E(Y) \geq E(X) - \epsilon.$$

Dado que $\epsilon > 0$ es arbitrario, se concluye que $\lim_{n \rightarrow \infty} E(X_n) \geq E(X)$. \square

El siguiente resultado establece otro tipo de condición suficiente para obtener la misma conclusión.

TEOREMA DE CONVERGENCIA DOMINADA. Sea X_1, X_2, \dots una sucesión de variables aleatorias para la cual existe otra variable Y integrable tal que $|X_n| \leq Y$, para $n \geq 1$. Si $\lim_{n \rightarrow \infty} X_n = X$ c.s., entonces X y X_n son integrables y

$$\lim_{n \rightarrow \infty} E(X_n) = E(X).$$

Demostración. Sea $Y_n = \inf\{X_n, X_{n+1}, \dots\}$. Entonces $Y_n \nearrow X$ cuando $n \rightarrow \infty$. Por lo tanto $(Y_n + Y) \nearrow (X + Y)$, en donde $Y_n + Y \geq 0$, pues como $-X_n \leq Y$, entonces $X_n \geq -Y$ para toda n , y por lo tanto $Y_n \geq -Y$. Por el teorema de convergencia monótona, $E(Y_n + Y) \nearrow E(X + Y)$. De donde se obtiene

$$E(Y_n) \nearrow E(X).$$

Sea ahora $Z_n = \sup\{X_n, X_{n+1}, \dots\}$. Entonces $Z_n \searrow X$ cuando $n \rightarrow \infty$. Por lo tanto $(Y - Z_n) \nearrow (Y - X)$, en donde $Y - Z_n \geq 0$, pues como $X_n \leq Y$ para toda n , entonces $Z_n \leq Y$. Por el teorema de convergencia monótona, $E(Y - Z_n) \nearrow E(Y - X)$. De donde se obtiene

$$E(Z_n) \searrow E(X).$$

Ahora observe que $Y_n \leq X_n \leq Z_n$. Por lo tanto $E(Y_n) \leq E(X_n) \leq E(Z_n)$. Al hacer n tender a infinito se obtiene el resultado. \square

Estos dos teoremas son herramientas fuertes en la teoría de la probabilidad. En particular, se usarán en la última parte del curso para formalizar algunas demostraciones.

7.4. Ejercicios

Convergencia casi segura

506. Para la convergencia casi segura se pide que el conjunto $\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}$ tenga probabilidad uno. Demuestre la medibilidad de tal conjunto probando que es idéntico al evento

$$\bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} (|X_n - X| \leq 1/k).$$

507. Demuestre que en la convergencia casi segura, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{c.s.} X$, y $X_n \xrightarrow{c.s.} Y$, entonces $X = Y$ casi seguramente. Sugerencia: $|X - Y| \leq |X - X_n| + |X_n - Y|$.
508. Demuestre que si $X_n \xrightarrow{c.s.} X$, entonces $aX_n + b \xrightarrow{c.s.} aX + b$, en donde a y b son constantes.
509. Demuestre que si $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$, entonces
- $X_n + Y_n \xrightarrow{c.s.} X + Y$.
 - $X_n Y_n \xrightarrow{c.s.} XY$.
510. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$, con P la medida de probabilidad uniforme. Demuestre que la sucesión $X_n = n1_{[0, 1/n]}$ converge casi seguramente a la variable aleatoria constante cero.
511. CONDICIÓN EQUIVALENTE PARA LA CONVERGENCIA CASI SEGURA. Demuestre que $X_n \xrightarrow{c.s.} X$ si, y sólo si, para cualquier $\epsilon > 0$,

$$P(|X_n - X| > \epsilon \text{ para una infinidad de valores de } n) = 0.$$

512. Use el ejercicio anterior para demostrar que si para cualquier $\epsilon > 0$, $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$, entonces $X_n \xrightarrow{c.s.} X$.

Convergencia en probabilidad

513. Demuestre que en la convergencia en probabilidad, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{p} X$, y $X_n \xrightarrow{p} Y$, entonces $X = Y$ casi seguramente.

Sugerencia: $P(|X - Y| > \epsilon) \leq P(|X - X_n| > \epsilon/2) + P(|X_n - Y| > \epsilon/2)$.

514. Considere el espacio de probabilidad $((0, 1], \mathcal{B}(0, 1], P)$, en donde P es la medida de probabilidad uniforme. Defina las variables aleatorias discretas

$$X_n = \sum_{k=1}^n \frac{k}{n} 1_{(\frac{k-1}{n}, \frac{k}{n}]}$$

Demuestre que X_n converge en probabilidad a una variable aleatoria con distribución uniforme en el intervalo $(0, 1]$.

515. Demuestre que si $X_n \xrightarrow{p} X$, entonces $aX_n + b \xrightarrow{p} aX + b$, en donde a y b son constantes.

516. Suponga que $X_n \xrightarrow{p} x$ y $Y_n \xrightarrow{p} y$, en donde x y y son dos números reales fijos. Demuestre que

a) $X_n + Y_n \xrightarrow{p} x + y$.

b) $X_n Y_n \xrightarrow{p} xy$.

c) Si g es continua en x , entonces $g(X_n) \xrightarrow{p} g(x)$.

517. Demuestre que si $X_n \xrightarrow{p} X$ y $Y_n \xrightarrow{p} Y$, entonces

a) $X_n + Y_n \xrightarrow{p} X + Y$.

b) $X_n Y_n \xrightarrow{p} XY$.

518. Sean X_1, X_2, \dots variables aleatorias independientes cada una con distribución unif $[a, b]$. Demuestre que cuando n tiende a infinito

a) $\min\{X_1, \dots, X_n\} \xrightarrow{p} a$.

b) $\max\{X_1, \dots, X_n\} \xrightarrow{p} b$.

519. Demuestre que si $X_n \xrightarrow{p} X$, entonces $X_n^2 \xrightarrow{p} X^2$.
520. Sea $c > 0$ una constante. Use la desigualdad de Chebyshev para demostrar que si X_n tiene distribución $\text{gama}(cn, n)$, entonces $X_n \xrightarrow{p} c$.

Convergencia en media

521. Demuestre que en la convergencia en media, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{m} X$, y $X_n \xrightarrow{m} Y$, entonces $X = Y$ casi seguramente. Sugerencia: $E|X - Y| \leq E|X - X_n| + E|X_n - Y|$.
522. Demuestre que si $X_n \xrightarrow{m} X$, entonces $aX_n + b \xrightarrow{m} aX + b$, en donde a y b constantes.
523. Suponga que $X_n \xrightarrow{m} X$ y $Y_n \xrightarrow{m} Y$. Demuestre que $X_n + Y_n \xrightarrow{m} X + Y$. Proporcione un contraejemplo para la afirmación: $X_n Y_n \xrightarrow{m} XY$.

Convergencia en media cuadrática

524. Demuestre que en la convergencia en media cuadrática, el límite es único casi seguramente, es decir, si $X_n \xrightarrow{m.c.} X$, y $X_n \xrightarrow{m.c.} Y$, entonces $X = Y$ casi seguramente. Sugerencia: Por la desigualdad c_r con $r = 2$, $E|X - Y|^2 \leq 2(E|X - X_n|^2 + E|X_n - Y|^2)$.
525. Demuestre que si $X_n \xrightarrow{m.c.} X$, entonces $aX_n + b \xrightarrow{m.c.} aX + b$, en donde a y b son constantes.
526. Use la desigualdad de Cauchy-Schwarz para demostrar que si $X_n \xrightarrow{m.c.} X$ y $Y_n \xrightarrow{m.c.} Y$, entonces $X_n + Y_n \xrightarrow{m.c.} X + Y$.

Convergencia en distribución

527. Demuestre que en la convergencia en distribución, el límite es único en distribución, es decir, si $X_n \xrightarrow{d} X$, y $X_n \xrightarrow{d} Y$, entonces X y Y

tienen la misma distribución.

Sugerencia: $|F_X(x) - F_Y(x)| \leq |F_X(x) - F_{X_n}(x)| + |F_{X_n}(x) - F_Y(x)|$.

528. Sea c una constante y suponga que $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$. Demuestre que

a) $cX_n \xrightarrow{d} cX$.

b) $X_n + c \xrightarrow{d} X + c$.

529. Demuestre que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$, entonces no necesariamente $X_n + Y_n \xrightarrow{d} X + Y$.

530. Demuestre que

a) si $X_n \xrightarrow{d} 0$, entonces $X_n \xrightarrow{p} 0$.

b) si $X_n \xrightarrow{d} 0$ y $Y_n \xrightarrow{d} 0$, entonces $X_n + Y_n \xrightarrow{d} 0$.

c) si $X_n \xrightarrow{d} 0$ y $Y_n \xrightarrow{d} 0$, entonces $X_n Y_n \xrightarrow{d} 0$.

531. Considere el espacio de probabilidad $([0, 1], \mathcal{B}[0, 1], P)$ en donde P es la medida de probabilidad uniforme. Demuestre que la sucesión $X_n = 1_{[0, 1/2+1/n]}$ converge en distribución a la variable aleatoria $X = 1_{[0, 1/2]}$.

532. Sea X_n con distribución unif $[a - 1/n, a + 1/n]$, en donde a es una constante. Demuestre que $X_n \xrightarrow{d} a$.

533. Sea X_n con distribución uniforme en el conjunto $\{0, 1, \dots, n\}$, y sea X continua con distribución uniforme en el intervalo $[0, 1]$. Demuestre que $\frac{1}{n}X_n \xrightarrow{d} X$.

534. Sea X con distribución uniforme en el conjunto $\{0, 1\}$. Demuestre que la siguiente sucesión de variables aleatorias converge en distribución pero no converge en probabilidad.

$$X_n = \begin{cases} X & \text{si } n \text{ es par,} \\ 1 - X & \text{si } n \text{ es impar.} \end{cases}$$

Relaciones entre los tipos de convergencia

535. OTRO EJEMPLO DE QUE LA CONV. CASI SEGURA NO IMPLICA LA CONV. EN MEDIA. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que para cada número natural n , $P(X_n = 0) = 1/4$, $P(X_n = 1) = 1/2$ y $P(X_n = 2) = 1/4$. Defina el producto $Y_n = X_1 X_2 \cdots X_n$. Demuestre que Y_n converge a cero, casi seguramente, pero no así en media, ni en media cuadrática.
536. Sea A_1, A_2, \dots una sucesión de eventos convergente al evento A . ¿En qué sentido la sucesión de variables aleatorias 1_{A_n} converge a 1_A ?
537. Sea X_n con distribución $N(\mu_n, \sigma_n^2)$ y X con distribución $N(\mu, \sigma^2)$. Suponga $\mu_n \rightarrow \mu$ y $\sigma_n^2 \rightarrow \sigma^2$, con $\sigma_n^2, \sigma^2 > 0$. ¿En qué sentido $X_n \rightarrow X$?

CAPÍTULO 8

Funciones generadoras

En este capítulo se estudia la función generadora de probabilidad, la función generadora de momentos y la función característica. Estas funciones son transformaciones de las distribuciones de probabilidad, y constituyen una herramienta muy útil en la teoría moderna de la probabilidad.

8.1. Función generadora de probabilidad

DEFINICIÓN. (FUNCIÓN GENERADORA DE PROBABILIDAD). La función generadora de probabilidad de una variable aleatoria X es la función

$$G(t) = E(t^X),$$

definida para valores reales de t tal que la esperanza sea convergente absolutamente.

Cuando sea necesario especificarlo se escribe $G_X(t)$ en lugar de $G(t)$, y se usan las letras f.g.p. en lugar de *función generadora de probabilidad*. Esta función se utiliza principalmente, aunque no únicamente, para variables

aleatorias con valores enteros. Supondremos tal caso y sin pérdida de generalidad consideraremos que las variables toman valores en el conjunto $\{0, 1, \dots\}$, que corresponde a la mayoría de las variables aleatorias discretas estudiadas en este curso. En tal situación,

$$G(t) = \sum_{k=0}^{\infty} t^k P(X = k).$$

Es decir, la f.g.p. es una serie de potencias en t , con coeficientes dados por la distribución de probabilidad, por ende el nombre de dicha función. Es importante observar que el radio de convergencia de esta serie es por lo menos uno, pues para $|t| < 1$,

$$|G(t)| \leq \sum_{k=0}^{\infty} |t|^k P(X = k) \leq \sum_{k=0}^{\infty} P(X = k) = 1.$$

Calculando la k -ésima derivada puede comprobarse además que a partir de la f.g.p. puede reconstruirse la función de densidad a través de la fórmula $P(X = k) = G^{(k)}(0)/k!$

EJEMPLO. Sea X con distribución Poisson(λ). La f.g.p. de X está definida para todo valor real de t y puede calcularse de la siguiente forma.

$$G(t) = \sum_{k=0}^{\infty} t^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{-\lambda(1-t)}.$$

▪

En la siguiente tabla se muestran ejemplos de funciones generadoras de probabilidad para algunas distribuciones discretas.

DISTRIBUCIÓN	FUNCIÓN GENERADORA DE PROBABILIDAD
$\text{unif}\{x_1, \dots, x_n\}$	$G(t) = (t^{x_1} + \dots + t^{x_n})/n$
$\text{Ber}(p)$	$G(t) = 1 - p + pt$
$\text{bin}(n, p)$	$G(t) = (1 - p + pt)^n$
$\text{geo}(p)$	$G(t) = p/[1 - t(1 - p)]$
$\text{Poisson}(\lambda)$	$G(t) = e^{-\lambda(1-t)}$
$\text{bin neg}(r, p)$	$G(t) = (p/[1 - t(1 - p)])^r$

La función generadora de probabilidad determina de manera única a la distribución en el siguiente sentido. Si X y Y tienen la misma distribución de probabilidad, entonces naturalmente $G_X(t) = G_Y(t)$, para valores de t donde esta esperanza exista. Inversamente, sean X y Y tales que $G_X(t)$ y $G_Y(t)$ existen y coinciden en algún intervalo no trivial alrededor del cero, entonces X y Y tienen la misma distribución. Estas y otras propiedades generales de la f.g.p. se estudian a continuación, más adelante se ilustran estos resultados con algunos ejemplos.

PROPOSICIÓN. (PROPIEDADES DE LA F.G.P.).

1. Sean X y Y variables aleatorias con valores en $\{0, 1, \dots\}$ tales que $G_X(t)$ y $G_Y(t)$ existen y coinciden en algún intervalo alrededor de $t = 0$. Entonces X y Y tienen la misma distribución de probabilidad.

2. Si el n -ésimo momento factorial de X existe, entonces

$$\lim_{t \nearrow 1} \frac{d^n}{dt^n} G_X(t) = E[X(X-1)\cdots(X-n+1)].$$

3. Sean X y Y independientes con f.g.p. $G_X(t)$ y $G_Y(t)$ respectivamente, entonces $G_{X+Y}(t) = G_X(t)G_Y(t)$.

Demostración.

1. Para cada $k \geq 0$, sean $a_k = P(X = k)$ y $b_k = P(Y = k)$. La igualdad $G_X(t) = G_Y(t)$ se escribe de la forma:

$$\sum_{k=0}^{\infty} t^k a_k = \sum_{k=0}^{\infty} t^k b_k.$$

Para que estas dos series de potencias en t coincidan en algún intervalo no trivial alrededor del cero, sus coeficientes deben forzosamente coincidir, es decir, $a_k = b_k$ para cada $k \geq 0$. Esto significa que las distribuciones de probabilidad coinciden.

2. Como las series de potencia se pueden derivar término a término con-

servándose el mismo radio de convergencia, se tiene que

$$\begin{aligned} G'(t) &= \frac{d}{dt} \sum_{k=0}^{\infty} t^k P(X = k) \\ &= \sum_{k=0}^{\infty} \frac{d}{dt} t^k P(X = k) \\ &= \sum_{k=1}^{\infty} k t^{k-1} P(X = k). \end{aligned}$$

Como por hipótesis la esperanza existe, por el lema de Abel (ver apéndice),

$$\lim_{t \nearrow 1} G'(t) = \sum_{k=1}^{\infty} k P(X = k) = E(X).$$

Para la segunda derivada se tiene

$$G''(t) = \sum_{k=2}^{\infty} k(k-1)t^{k-2} P(X = k),$$

de modo que cuando el segundo momento existe,

$$\lim_{t \nearrow 1} G''(t) = \sum_{k=2}^{\infty} k(k-1)P(X = k) = E(X(X-1)).$$

De manera análoga se demuestra para las derivadas de orden superior.

3. Cuando X y Y son independientes,

$$G_{X+Y}(t) = E(t^{X+Y}) = E(t^X t^Y) = E(t^X) E(t^Y) = G_X(t) G_Y(t).$$

□

EJEMPLO. Se ha encontrado que la f.g.p. de una variable aleatoria X con distribución Poisson(λ) es $G(t) = e^{-\lambda(1-t)}$. Usando esta función encontraremos la esperanza y varianza de X . Al derivar una vez se obtiene $G'(t) =$

$\lambda e^{-\lambda(1-t)}$, y al evaluar en $t = 1$, $E(X) = G'(1) = \lambda$. Derivando por segunda vez, $G''(t) = \lambda^2 e^{-\lambda(1-t)}$, y en $t = 1$ se obtiene $E(X(X-1)) = G''(1) = \lambda^2$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$. ■

Debido a la segunda propiedad, a la f.g.p. también se le conoce como *función generadora de momentos factoriales*. Ahora se muestra el uso de esta función para determinar la distribución de una variable aleatoria, el procedimiento es elegante y sencillo.

EJEMPLO. Suponga que X y Y son independientes con distribución Poisson(λ_1) y Poisson(λ_2), respectivamente. Entonces

$$G_{X+Y}(t) = G_X(t) G_Y(t) = e^{-\lambda_1(1-t)} e^{-\lambda_2(1-t)} = e^{-(\lambda_1+\lambda_2)(1-t)}.$$

Esta expresión corresponde a la f.g.p. de la distribución Poisson con parámetro $\lambda_1 + \lambda_2$. Debido a la unicidad, $X + Y$ tiene distribución Poisson($\lambda_1 + \lambda_2$). ■

La definición de función generadora de probabilidad puede extenderse al caso de vectores aleatorios de la siguiente forma. La f.g.p. del vector (X, Y) es la función $G_{X,Y}(s, t) = E(s^X t^Y)$, para valores reales de s y t donde esta esperanza sea absolutamente convergente. Puede demostrarse que las variables X y Y son independientes si, y sólo si, $G_{X,Y}(s, t) = G_X(s) G_Y(t)$. La definición de f.g.p. para vectores de dimensión mayor es análoga.

8.2. Función generadora de momentos

Esta es otra función que se puede asociar a algunas distribuciones de probabilidad. Su existencia no está garantizada en todos los casos, pero cuando existe, determina de manera única a la distribución de probabilidad asociada, y tiene propiedades semejantes a las de la función generadora de probabilidad. La función generadora de momentos se utiliza tanto para variables aleatorias discretas como continuas.

DEFINICIÓN. (FUNCIÓN GENERADORA DE MOMENTOS). La función generadora de momentos de la variable aleatoria X es la función

$$M(t) = E(e^{tX}),$$

definida para valores reales de t tales que la esperanza es absolutamente convergente.

Nuevamente, cuando sea necesario especificarlo se escribe $M_X(t)$ en lugar de $M(t)$, y se usan las letras f.g.m. en lugar del término *función generadora de momentos*. La parte importante de esta función es su existencia en una vecindad no trivial alrededor del cero. Observe que la f.g.m. y la f.g.p. están relacionadas, cuando existen, por la igualdad $M(t) = G(e^t)$.

EJEMPLO. Sea X con distribución gama(n, λ). Entonces la f.g.m. de X puede calcularse de la siguiente forma.

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} dx \\ &= \lambda^n (\lambda - t)^{-n} \int_0^\infty \frac{[(\lambda - t)x]^{n-1}}{\Gamma(n)} (\lambda - t) e^{-(\lambda - t)x} dx \\ &= [\lambda / (\lambda - t)]^n. \end{aligned}$$

La última integral vale uno pues el integrando es la función de densidad de una distribución gama. Observe que $M(t)$ esta definida únicamente para valores de t menores que λ . ▪

La siguiente tabla muestra algunos otros ejemplos de funciones generadoras de momentos para ciertas distribuciones continuas.

DISTRIBUCIÓN	FUNCIÓN GENERADORA DE MOMENTOS
$\text{unif}(a, b)$	$M(t) = (e^{bt} - e^{at})/(bt - at)$
$\text{exp}(\lambda)$	$M(t) = \lambda/(\lambda - t)$
$\text{gama}(n, \lambda)$	$M(t) = [\lambda/(\lambda - t)]^n$
$N(\mu, \sigma^2)$	$M(t) = \exp(\mu t + \sigma^2 t^2/2)$
$\chi^2(n)$	$M(t) = (1 - 2t)^{-n/2}$
$t(n)$	$M(t)$ no existe para $t \neq 0$

Se demuestran a continuación algunas propiedades básicas de la f.g.m., y después se muestra su utilidad mediante algunos ejemplos.

PROPOSICIÓN. Sea X con f.g.m. $M(t)$ finita para cada $t \in (-s, s)$, para algún $s > 0$. Entonces

1. Todos los momentos de X son finitos.

2.
$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n).$$

3. $M(t)$ tiene derivadas continuas de cualquier orden en $(-s, s)$, y se cumple

$$\left. \frac{d^n}{dt^n} M(t) \right|_{t=0} = E(X^n).$$

Demostración.

1. La prueba se basa en las identidades:

$$E|X|^n = n \int_0^\infty (1 - F(x)) x^{n-1} dx + n \int_{-\infty}^0 F(x) |x|^{n-1} dx,$$

y

$$M(t) = 1 + t \int_0^\infty (1 - F(x)) e^{tx} dx - t \int_{-\infty}^0 F(x) e^{tx} dx,$$

en donde, por hipótesis, las dos integrales de $M(t)$ son finitas para cualquier $t \in (-s, s)$. Demostraremos que cada integral de la expresión de $E|X|^n$ es menor o igual a la correspondiente integral de $M(t)$. Para el caso $x > 0$ se toma cualquier $t \in (0, s)$, y entonces

$$\frac{(tx)^n}{n!} \leq e^{tx}.$$

Es decir, $x^n \leq (n!/t^n)e^{tx}$. De modo que, salvo constantes, la primera integral de $E|X|^n$ es menor o igual a la primera integral de $M(t)$, siendo ésta última finita, la primera también. Para el caso $x < 0$ conviene tomar $t \in (-s, 0)$, pues en tal caso $tx > 0$ y entonces

$$\frac{|tx|^n}{n!} \leq e^{|tx|} = e^{tx}.$$

Es decir, $|x|^n \leq (n!/|t|^n)e^{tx}$. Ahora la segunda integral de $E|X|^n$ es menor o igual a la segunda integral de $M(t)$, siendo ésta última finita, la primera también. De esta forma todos los momentos de X existen cuando $M(t)$ es finita en algún intervalo no trivial alrededor del cero.

2. Se usa la fórmula

$$E(X^n) = n \int_0^\infty (1 - F(x)) x^{n-1} dx - n \int_{-\infty}^0 F(x) x^{n-1} dx.$$

Entonces para cualquier $t \in (-s, s)$, y $m \geq 1$,

$$\begin{aligned} \sum_{n=0}^m \frac{t^n}{n!} E(X^n) &= 1 + \sum_{n=1}^m \frac{t^n}{n!} n \int_0^{\infty} (1 - F(x)) x^{n-1} dx \\ &\quad - \sum_{n=1}^m \frac{t^n}{n!} n \int_{-\infty}^0 F(x) x^{n-1} dx \\ &= 1 + t \int_0^{\infty} (1 - F(x)) \sum_{n=0}^{m-1} \frac{t^n}{n!} x^n dx \\ &\quad - t \int_{-\infty}^0 F(x) \sum_{n=0}^{m-1} \frac{t^n}{n!} x^n dx. \end{aligned}$$

Usando el teorema de convergencia monótona, o el de convergencia dominada, dependiendo de los valores de t y x , cada una de estas integrales es convergente, para cualquier $t \in (-s, s)$, cuando se hace m tender a infinito. De modo que

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) &= 1 + t \int_0^{\infty} (1 - F(x)) e^{tx} dx - t \int_{-\infty}^0 F(x) e^{tx} dx \\ &= M(t). \end{aligned}$$

3. Dado que $M(t)$ se puede expresar como una serie de potencias en t , diferenciando y evaluando en cero se obtienen los coeficientes $E(X^n)$.

□

NOTA IMPORTANTE. El hecho de que el n -ésimo momento de una variable aleatoria exista, no implica que éste puede ser hallado a través de la n -ésima derivada de la f.g.m. evaluada en cero. Es decir, es necesario conocer la existencia de la f.g.m. para que pueda ser utilizada para obtener los momentos. Por ejemplo, una variable aleatoria con distribución $t(n)$ tiene esperanza cero pero su f.g.m. $M(t)$ no existe para t distinto de cero.

EJEMPLO. Sea X con distribución gama(n, λ). Hemos encontrado antes que para $t < \lambda$, $M(t) = \lambda^n (\lambda - t)^{-n}$. Calcularemos ahora la esperanza y varianza de X con ayuda de la f.g.m. Derivando una vez, $M'(t) = \lambda^n n (\lambda - t)^{-n-1}$. Al evaluar en $t = 0$ se obtiene $E(X) = n/\lambda$. Derivando nuevamente, $M''(t) = \lambda^n n(n+1) (\lambda - t)^{-n-2}$. Por lo tanto $E(X^2) = M''(0) = n(n+1)/\lambda^2$. Entonces $\text{Var}(X) = n(n+1)/\lambda^2 - n^2/\lambda^2 = n/\lambda^2$. ■

EJEMPLO. Suponga ahora que X y Y son independientes cada una con distribución gama(n, λ) y gama(m, λ), respectivamente. Entonces la f.g.m. de $X + Y$ es

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \lambda^n (\lambda - t)^{-n} \lambda^m (\lambda - t)^{-m} = \lambda^{n+m} (\lambda - t)^{-n-m}.$$

Esta es la expresión de la f.g.m. de la distribución gama, ahora con parámetros $n+m$ y λ . Se concluye entonces $X+Y$ tiene distribución gama($n+m, \lambda$). ■

Nuevamente, es sencillo demostrar que la función generadora de la suma de dos variables aleatorias independientes es el producto de las funciones generadoras individuales.

PROPOSICIÓN. Sean X y Y son independientes, y cuyas f.g.m. existen en una vecindad no trivial alrededor del cero. Entonces para cualquier $t \in (-s, s)$ para algún $s > 0$,

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Demostración.

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) = M_X(t) M_Y(t).$$

□

Es interesante observar que la condición $M_{X+Y}(t) = M_X(t) M_Y(t)$ no es suficiente para concluir que X y Y son independientes.

EJERCICIO. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = [1 + xy(x^2 - y^2)]/4, \quad \text{para } -1 < x, y < 1.$$

Demuestre que X y Y no son independientes y sin embargo se cumple la identidad $M_{X+Y}(t) = M_X(t) M_Y(t)$. ■

Como hemos mencionado antes, no todas las distribuciones de probabilidad permiten calcular la función generadora de momentos dentro de un intervalo no trivial alrededor del cero, ni todos los cálculos son tan sencillos como en el ejemplo mostrado. Por ejemplo, la f.g.m. de la distribución Cauchy estándar no existe para valores de t distintos de cero, esto se pide comprobar en el ejercicio 576. Por otro lado, cuando se tienen dos variables X y Y con la misma distribución, entonces sus funciones generadoras de momentos coinciden pues éstas se obtienen a través de la función de distribución común. Por el contrario, si $M_X(t) = M_Y(t)$ en una vecindad no trivial alrededor del cero, entonces puede demostrarse que sus distribuciones coinciden, este resultado y otro relativo a convergencia es el contenido de la siguiente proposición, cuya demostración omitiremos.

PROPOSICIÓN.

1. (UNICIDAD). Las variables X y Y tienen la misma distribución si, y sólo si, $M_X(t) = M_Y(t)$ para valores de t en una vecindad no trivial alrededor del cero.
2. (CONTINUIDAD). Sea X_1, X_2, \dots una sucesión de variables aleatorias cuyas funciones generadoras de momentos existen todas ellas en algún intervalo no trivial alrededor del cero. Sea X con f.g.m. $M_X(t)$. Entonces $X_n \xrightarrow{d} X$ si, y sólo si, $M_{X_n}(t) \rightarrow M_X(t)$.

Para el caso de vectores aleatorios se tiene la siguiente definición. La fun-

ción generadora de momentos del vector (X, Y) es la función $M_{X,Y}(s, t) = E(e^{sX} e^{tY})$, para valores reales de s y t donde esta esperanza sea absolutamente convergente. Puede demostrarse que las variables X y Y son independientes si, y sólo si, $M_{X,Y}(s, t) = M_X(s) M_Y(t)$. La definición de f.g.m. para vectores de dimensión mayor es análoga.

En la sección de ejercicios se pueden encontrar las funciones generadoras de momentos de algunas otras distribuciones de probabilidad, tanto discretas como continuas, así como en el primer apéndice al final del libro.

8.3. Función característica

Esta es una función definida para cada distribución de probabilidad, y a diferencia de las funciones generadoras de probabilidad y de momentos estudiadas antes, siempre existe.

DEFINICIÓN. (FUNCIÓN CARACTERÍSTICA). La función característica de la variable aleatoria X es la función

$$\phi(t) = E(e^{itX}),$$

definida para cualquier número real t . El número i es la unidad de los números imaginarios.

Observe que la transformación $X \mapsto e^{itX}$ lleva una variable aleatoria real X a una variable aleatoria con valores en los números complejos de la forma $\cos(tX) + i \sin(tX)$, en donde cada parte de este número complejo es una variable aleatoria real, es decir, se trata de un vector aleatorio bidimensional como los estudiados anteriormente. La función característica puede entonces escribirse en la forma

$$\phi(t) = E(\cos tX) + i E(\sin tX).$$

Nuevamente se escribe $\phi_X(t)$ cuando sea necesario especificar que se trata de

la función característica de X , y se escribe simplemente f.c. en lugar de *función característica*. Observe que la f.c., la f.g.m. y la f.g.p. están relacionadas, cuando existen las dos últimas, por las igualdades $\phi(t) = M(it) = G(e^{it})$. Se muestran a continuación algunos ejemplos de la forma de encontrar la función característica a partir de una distribución de probabilidad.

EJEMPLO. Sea X con distribución $\text{bin}(n, p)$. Entonces

$$\begin{aligned}\phi(t) &= E(e^{itX}) \\ &= \sum_{x=0}^n e^{itx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^{it})^x (1-p)^{n-x} \\ &= (1-p + pe^{it})^n.\end{aligned}$$

EJEMPLO. Sea X con distribución $\text{Poisson}(\lambda)$. Entonces

$$\begin{aligned}\phi(t) &= E(e^{itX}) \\ &= \sum_{x=0}^{\infty} e^{itx} e^{-\lambda} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!} \\ &= e^{-\lambda(1-e^{it})}.\end{aligned}$$

Otros ejemplos de funciones características de distribuciones discretas se muestra en la siguiente tabla. El lector puede comprobar cada una de estas expresiones.

DISTRIBUCIÓN	FUNCIÓN CARACTERÍSTICA
Ber(p)	$\phi(t) = 1 - p + pe^{it}$
bin(n, p)	$\phi(t) = (1 - p + pe^{it})^n$
Poisson(λ)	$\phi(t) = e^{-\lambda(1-e^{it})}$
geo(p)	$\phi(t) = p/(1 - (1 - p)e^{it})$
bin neg(r, p)	$\phi(t) = [p/(1 - (1 - p)e^{it})]^r$

Ahora se mostrará la forma de encontrar la función característica para dos distribuciones continuas: la distribución normal y la distribución gama.

EJEMPLO. Sea X con distribución $N(\mu, \sigma^2)$. Entonces

$$\begin{aligned}
 \phi(t) &= E(e^{itX}) \\
 &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2-2x(\mu-it\sigma^2)+\mu^2)/2\sigma^2} dx \\
 &= e^{(-\mu^2+(\mu-it\sigma^2)^2)/2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[x-(\mu-it\sigma^2)]^2/2\sigma^2} dx \\
 &= e^{it\mu-t^2\sigma^2/2}.
 \end{aligned}$$

Observe que el último integrando es la función de densidad normal con media el número complejo $\mu - it\sigma^2$, y varianza σ^2 . El hecho de que esta integral también vale uno puede comprobarse, por ejemplo, usando el *principio de*

continuación analítica de la teoría de variable compleja. ■

EJEMPLO. Sea X con distribución gama(n, λ). Entonces

$$\begin{aligned}
 \phi(t) &= E(e^{itX}) \\
 &= \int_0^\infty e^{itx} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} dx \\
 &= \int_0^\infty \frac{\lambda}{\Gamma(n)} (\lambda x)^{n-1} e^{-(\lambda-it)x} dx \\
 &= \frac{\lambda^n}{(\lambda-it)^n} \int_0^\infty \frac{[(\lambda-it)x]^{n-1}}{\Gamma(n)} (\lambda-it) e^{-(\lambda-it)x} dx \\
 &= \left(\frac{\lambda}{\lambda-it}\right)^n.
 \end{aligned}$$

El último integrando es la función de densidad de la distribución gama($n, \lambda-it$). Nuevamente usando la teoría de variable compleja puede demostrarse rigurosamente que esta integral también vale uno. ■

La siguiente tabla muestra algunos otros ejemplos de funciones características para variables aleatorias continuas.

DISTRIBUCIÓN	FUNCIÓN CARACTERÍSTICA
$\text{unif}(a, b)$	$\phi(t) = (e^{ibt} - e^{iat}) / (ibt - iat)$
$\text{exp}(\lambda)$	$\phi(t) = \lambda / (\lambda - it)$
$\text{gama}(n, \lambda)$	$\phi(t) = [\lambda / (\lambda - it)]^n$
$N(\mu, \sigma^2)$	$\phi(t) = \exp(i\mu t - \sigma^2 t^2 / 2)$
$\chi^2(n)$	$\phi(t) = (1 - 2it)^{-n/2}$
$t(n)$	$\phi(t) = e^{- t }$, cuando $n = 1$.

La existencia de la función característica para cualquier distribución de probabilidad se sigue del siguiente resultado.

PROPOSICIÓN. (EXISTENCIA). Para cualquier número real t , $|\phi(t)| \leq 1$. En particular, $\phi(0) = 1$.

Demostración. Para cualquier número real t ,

$$|\phi(t)| = \left| \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |e^{itx}| dF(x) = \int_{-\infty}^{\infty} dF(x) = 1.$$

□

De modo que $\phi(t)$ es un número complejo de módulo menor o igual a uno, para cualquier valor de t . Veremos a continuación algunas otras propiedades de esta importante función. En particular, demostraremos que los momentos de una variable aleatoria X pueden ser generados, cuando existen, con la f.c. a través de la fórmula $\phi^{(n)}(0) = i^n E(X^n)$, y como en el caso de las funciones generadoras anteriores, cuando X y Y son independientes se cumple que $\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$, no siendo válido en general el recíproco.

PROPOSICIÓN. Si X tiene n -ésimo momento finito, entonces

1. $\frac{d^n}{dt^n} \phi(t) \Big|_{t=0} = i^n E(X^n)$.

2. Cuando $t \rightarrow 0$,

$$\phi(t) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} E(X^k) + \frac{(it)^n}{n!} (E(X^n) + o(1)). \quad (8.1)$$

Demostración.

1. Para cualquier h distinto de cero,

$$\begin{aligned}
 \frac{\phi(t+h) - \phi(t)}{h} &= \int_{-\infty}^{\infty} \frac{e^{i(t+h)x} - e^{itx}}{h} dF(x) \\
 &= \int_{-\infty}^{\infty} e^{itx} \frac{e^{ihx} - 1}{h} dF(x) \\
 &= E\left(e^{itX} \frac{e^{ihX} - 1}{h} \right). \tag{8.2}
 \end{aligned}$$

Como $\lim_{h \rightarrow 0} \frac{e^{ihx} - 1}{h} = ix$, entonces, puntualmente,

$$\lim_{h \rightarrow 0} e^{itX} \frac{e^{ihX} - 1}{h} = iX e^{itX}.$$

Comprobaremos que las variables aleatorias de esta sucesión, parametrizada por h , están uniformemente acotadas por una variable aleatoria integrable, en efecto,

$$\begin{aligned}
 \left| e^{itX} \frac{e^{ihX} - 1}{h} \right| &= \left| \frac{e^{ihX} - 1}{h} \right| \\
 &= \left| \frac{1}{h} \int_0^h iX e^{isX} ds \right| \\
 &\leq |X| \frac{1}{h} \int_0^h |e^{isX}| ds \\
 &= |X|.
 \end{aligned}$$

Por hipótesis, $E|X| < \infty$, de modo que usando el teorema de convergencia dominada en (8.2) se obtiene

$$\frac{d}{dt} \phi(t) = E[iX e^{itX}].$$

Por el mismo procedimiento se encuentra que

$$\frac{d^n}{dt^n} \phi(t) = E[(iX)^n e^{itX}].$$

Tomando el límite cuando $t \rightarrow 0$ y usando nuevamente el teorema de convergencia dominada, se demuestra finalmente que

$$\left. \frac{d^n}{dt^n} \phi(t) \right|_{t=0} = i^n E(X^n).$$

2. La fórmula se sigue del inciso anterior y del siguiente resultado de análisis. Si g es una función con valores reales o complejos y definida en algún intervalo no trivial alrededor del origen con $g^{(n)}(0)$ finita, entonces cuando $t \rightarrow 0$,

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2!}g''(0) + \cdots + \frac{t^{n-1}}{(n-1)!}g^{(n-1)}(0) + \frac{t^n}{n!}(g^{(n)}(0) + o(1)).$$

□

En la última parte del curso se usará la expansión (8.1) para demostrar la ley de los grandes números y el teorema del límite central. Para el primer resultado se supondrá el primer momento finito y la expansión adquiere la expresión $\phi(t) = 1 + it(E(X) + o(1))$, cuando $t \rightarrow 0$. Para el teorema del límite central se supondrá el segundo momento finito y la expresión que se usa es $\phi(t) = 1 + itE(X) + ((it)^2/2!)(E(X^2) + o(1))$, cuando $t \rightarrow 0$.

PROPOSICIÓN. Si X y Y son independientes, entonces $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

Demostración. Por independencia,

$$\phi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX} e^{itY}) = E(e^{itX}) E(e^{itY}) = \phi_X(t)\phi_Y(t).$$

□

NOTA IMPORTANTE. El resultado anterior establece en particular que el producto de dos funciones características es nuevamente una función característica. Por otro lado, es necesario señalar que la condición $\phi_{X+Y}(t) =$

$\phi_X(t)\phi_Y(t)$ no es suficiente para concluir que las variables aleatorias X y Y son independientes.

EJERCICIO. Sea (X, Y) un vector aleatorio con función de densidad

$$f(x, y) = [1 + xy(x^2 - y^2)]/4, \quad \text{para } -1 < x, y < 1.$$

Demuestre que X y Y no son independientes y sin embargo se cumple la identidad $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$. \bullet

Otra de las propiedades fundamentales de la función característica es su capacidad de determinar de manera única a las distribuciones de probabilidad. A este respecto se tienen los siguientes resultados.

PROPOSICIÓN. (FÓRMULA DE INVERSIÓN DE LÈVY). Sea X con función de distribución $F(x)$, y función característica $\phi(t)$. Si $x < y$ son puntos de continuidad de F , entonces

$$F(y) - F(x) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt.$$

Cuando x y y no necesariamente son puntos de continuidad de F , el lado izquierdo es $\frac{1}{2}(F(y) + F(y-)) - \frac{1}{2}(F(x) + F(x-))$.

Demostración. Para $T > 0$ sea

$$\begin{aligned} I(T) &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt \\ &= \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \left[\int_{-\infty}^{\infty} e^{itz} dF(z) \right] dt \\ &= \frac{1}{2\pi} \int_{-T}^T \int_{-\infty}^{\infty} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dF(z) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{e^{it(z-x)} - e^{it(z-y)}}{it} dt dF(z). \end{aligned}$$

El cambio en el orden de integración es permitido pues el integrando es una función continua y acotada en $t \in [-T, T]$ y $z \in \mathbb{R}$, incluyendo cuando $t = 0$, pues puede definirse esta función de acuerdo a su comportamiento límite en ese punto, es decir,

$$\lim_{t \rightarrow 0} \frac{e^{it(z-x)} - e^{it(z-y)}}{it} = y - x.$$

Desarrollando las exponenciales en términos de senos y cosenos se obtiene

$$I(T) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-T}^T \frac{1}{it} (\cos t(z-x) + i \operatorname{sen} t(z-x) - \cos t(z-y) - i \operatorname{sen} t(z-y)) dt dF(z),$$

en donde para cualquier número real a , por ser coseno una función par, y seno una función impar,

$$\int_{-T}^T \frac{\cos(at)}{t} dt = 0,$$

y

$$\int_{-T}^T \frac{\operatorname{sen}(at)}{t} dt = 2 \int_0^T \frac{\operatorname{sen}(at)}{t} dt.$$

Por lo tanto

$$I(T) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (2 \int_0^T \frac{\operatorname{sen} t(z-x)}{t} dt - 2 \int_0^T \frac{\operatorname{sen} t(z-y)}{t} dt) dF(z).$$

El siguiente paso consiste en aplicar el teorema de convergencia dominada cuando $T \rightarrow \infty$. La integral $I(T)$ es la esperanza de la variable aleatoria

$$X_T = \frac{1}{2\pi} (2 \int_0^T \frac{\operatorname{sen} t(X-x)}{t} dt - 2 \int_0^T \frac{\operatorname{sen} t(X-y)}{t} dt).$$

Nos interesa encontrar el límite de esta variable cuando $T \rightarrow \infty$. Para ello se hace uso del siguiente resultado no trivial:

$$\lim_{T \rightarrow \infty} 2 \int_0^T \frac{\operatorname{sen} at}{t} dt = \pi \operatorname{signo}(a) = \begin{cases} \pi & \text{si } a > 0, \\ -\pi & \text{si } a < 0, \\ 0 & \text{si } a = 0. \end{cases}$$

Entonces, puntualmente,

$$\begin{aligned}
 \lim_{T \rightarrow \infty} X_T &= \frac{1}{2\pi} (\pi \operatorname{signo}(X - x) - \pi \operatorname{signo}(X - y)) \\
 &= \frac{1}{2} 1_{\{x,y\}}(X) + 1_{(x,y)}(X) \\
 &= \begin{cases} 0 & \text{si } X < x, \\ 1/2 & \text{si } X = x, \\ 1 & \text{si } x < X < y, \\ 1/2 & \text{si } X = y, \\ 0 & \text{si } X > y. \end{cases}
 \end{aligned}$$

Además, las variables X_T están acotadas en valor absoluto por una constante pues para cualquier número real a ,

$$\left| \int_0^T \frac{\operatorname{sen} at}{t} dt \right| \leq \sup_{T>0} \left| \int_0^T \frac{\operatorname{sen} t}{t} dt \right| < \infty.$$

Por lo tanto

$$\begin{aligned}
 \lim_{T \rightarrow \infty} I(T) &= \int_{-\infty}^{\infty} \left[\frac{1}{2} 1_{\{x,y\}}(z) + 1_{(x,y)}(z) \right] dF(z) \\
 &= \frac{1}{2} P(X = x) + \frac{1}{2} P(X = y) + P(x < X < y) \\
 &= P(x < X \leq y) + \frac{1}{2} P(X = x) - \frac{1}{2} P(X = y) \\
 &= F(y) - F(x) + \frac{1}{2} P(X = x) - \frac{1}{2} P(X = y) \\
 &= \frac{1}{2} (F(y) + F(y-)) - \frac{1}{2} (F(x) + F(x-)).
 \end{aligned}$$

En particular, si x y y son puntos de continuidad de F , entonces el límite de la integral es igual a $F(y) - F(x)$. \square

Como corolario del teorema de inversión demostraremos que la función característica determina de manera única a la distribución de probabilidad.

TEOREMA DE UNICIDAD. Si X y Y son tales que $\phi_X(t) = \phi_Y(t)$ para todo valor real de t , entonces X y Y tienen la misma distribución.

Demostración. Sea $\phi(t)$ la función característica común. Sea z cualquier número real, y sean x y y tales que $x < z < y$. Haciendo x tender a $-\infty$, y $y \searrow z$, en la fórmula de inversión de Lèvy, se obtiene una única función de distribución dada por

$$F(z) = \lim_{y \searrow z} \lim_{x \searrow -\infty} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt.$$

□

Cuando la condición $\phi_X(t) = \phi_Y(t)$ sólo se cumple en una vecindad del cero, no es necesariamente cierto que la distribución de probabilidad queda completamente especificada. Véase [13] para un ejemplo al respecto.

En el caso absolutamente continuo se tiene la siguiente fórmula explícita.

PROPOSICIÓN (FÓRMULA DE INVERSIÓN EN EL CASO ABS. CONTINUO). Sea X absolutamente continua con función de densidad $f(x)$, y función característica $\phi(t)$. Entonces

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt.$$

Demostración. Sean $x < y$, dos puntos de continuidad de F . Por el teorema

de inversión de Lèvy, y por el teorema de Fubini,

$$\begin{aligned}
 F(y) - F(x) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_x^y e^{-itx} dx \right] \phi(t) dt. \\
 &= \int_x^y \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt \right] dx.
 \end{aligned}$$

Por lo tanto el integrando debe ser la función de densidad de X . \square

Es necesario señalar que el uso de esta fórmula requiere conocer de antemano que la función característica proviene de una variable aleatoria absolutamente continua. De aquí surge el problema, que únicamente mencionamos, de encontrar condiciones sobre $\phi(t)$ que garanticen que la correspondiente variable aleatoria es absolutamente continua.

Ahora se demuestra un resultado que será de utilidad en la última parte del curso y que establece que la convergencia en distribución es equivalente a la convergencia puntual de las correspondientes funciones características. El resultado es válido como esta enunciado pero sólo demostraremos una de las implicaciones.

TEOREMA DE CONTINUIDAD. Sean X, X_1, X_2, \dots variables aleatorias. Entonces $X_n \xrightarrow{d} X$ si, y sólo si, $\phi_{X_n}(t) \rightarrow \phi_X(t)$.

Demostración. (\Leftarrow) Suponga que $\phi_{X_n}(t) \rightarrow \phi_X(t)$. Entonces para dos puntos de continuidad $x < y$ de F_X , el teorema de inversión de Lèvy establece

que

$$\begin{aligned}
 F_X(y) - F_X(x) &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \phi(t) dt. \\
 &= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} \left[\lim_{n \rightarrow \infty} \phi_{X_n}(t) \right] dt. \\
 &= \lim_{n \rightarrow \infty} \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx} - e^{-ity}}{it} [\phi_{X_n}(t)] dt. \\
 &= \lim_{n \rightarrow \infty} F_{X_n}(y) - F_{X_n}(x).
 \end{aligned}$$

Haciendo x tender a $-\infty$ se obtiene $F_X(y) = \lim_{n \rightarrow \infty} F_{X_n}(y)$. \square

En el siguiente capítulo usaremos este resultado para demostrar el teorema central del límite. Finalmente mencionamos la definición de función característica para vectores aleatorios. La f.c. del vector (X, Y) es la función $\phi_{X,Y}(s, t) = E(e^{isX} e^{itY})$, para valores reales de s y t donde esta esperanza sea absolutamente convergente. Nuevamente puede demostrarse que las variables X y Y son independientes si, y sólo si, $\phi_{X,Y}(s, t) = \phi_X(s) \phi_Y(t)$. De manera análoga puede definirse la función característica para vectores de dimensión mayor.

8.4. Ejercicios

Función generadora de probabilidad

538. Sea X con varianza finita y con f.g.p. $G(t)$. Demuestre que
- $E(X) = G'(1-)$.
 - $E(X^2) = G''(1-) + G'(1-)$.
 - $\text{Var}(X) = G''(1-) + G'(1-) - [G'(1-)]^2$.
539. Sean X y Y independientes, y sean a y b dos constantes. Demuestre que
- $P(X = k) = G^{(k)}(0)/k!$ para $k = 0, 1, \dots$
 - $G_{aX+b}(t) = t^b G_X(t^a)$.
 - $G_{X-Y}(t) = G_X(t) G_Y(1/t)$.
540. Sean X_1, \dots, X_n independientes tales que X_k tiene f.g.p. $G_k(t)$, para $k = 1, \dots, n$. Demuestre que $G_{X_1+\dots+X_n}(t) = G_1(t) \cdots G_n(t)$.
541. Demuestre o proporcione un contraejemplo: Si $G_{X+Y}(t) = G_X(t) \cdot G_Y(t)$, para valores de t en algún intervalo no trivial alrededor del cero, entonces X y Y son independientes.
542. Sea X_1, X_2, \dots una sucesión de v.a.i.i.d. con f.g.p. $G_X(t)$. Sea N otra variable aleatoria con valores en \mathbb{N} , independiente de la sucesión y con f.g.p. $G_N(t)$. Sea $S = X_1 + \dots + X_N$. Demuestre que
- $G_S(t) = G_N(G_X(t))$.
 - $E(S) = E(N)E(X)$, usando $G_S(t)$.
 - $\text{Var}(S) = E^2(X) \text{Var}(N) + E(N) \text{Var}(X)$, usando $G_S(t)$.
543. Encuentre la función generadora de probabilidad, si existe, de una variable aleatoria con función de densidad

$$a) f(x) = \frac{1}{x!(e-1)}, \text{ para } x = 1, 2, \dots$$

$$b) f(x) = \frac{1}{x(x+1)}, \text{ para } x = 1, 2, \dots$$

544. Sea X con distribución $\text{Ber}(p)$. Demuestre que

$$a) G(t) = 1 - p + pt.$$

$$b) E(X) = p, \text{ usando } G(t).$$

$$c) \text{Var}(X) = p(1-p), \text{ usando } G(t).$$

$$d) E(X^n) = p, \text{ usando } G(t).$$

545. Sea X con distribución $\text{bin}(n, p)$. Demuestre que

$$a) G(t) = (1 - p + pt)^n.$$

$$b) E(X) = np, \text{ usando } G(t).$$

$$c) \text{Var}(X) = np(1-p), \text{ usando } G(t).$$

546. Sean X_1, \dots, X_n variables aleatorias independientes, cada una con distribución $\text{Ber}(p)$. Use la f.g.p. para demostrar que la variable $X_1 + \dots + X_n$ tiene distribución $\text{bin}(n, p)$.

547. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$, respectivamente. Use la f.g.p. para demostrar que la variable $X + Y$ tiene distribución $\text{bin}(n + m, p)$.

548. Sea X con distribución $\text{bin}(N, p)$, en donde N es una variable aleatoria con distribución $\text{bin}(n, r)$. Use la f.g.p. para demostrar que X tiene distribución $\text{bin}(n, rp)$.

549. Sea X con distribución $\text{geo}(p)$. Demuestre que

$$a) G(t) = p/[1 - t(1-p)].$$

$$b) E(X) = (1-p)/p, \text{ usando } G(t).$$

$$c) \text{Var}(X) = (1-p)/p^2, \text{ usando } G(t).$$

550. Sea X con distribución Poisson(λ). Demuestre que

a) $G(t) = e^{-\lambda(1-t)}$.

b) $E(X) = \lambda$, usando $G(t)$.

c) $\text{Var}(X) = \lambda$, usando $G(t)$.

551. Sean X y Y independientes con distribución Poisson con parámetros λ_1 y λ_2 respectivamente. Use la f.g.p. para demostrar que la variable $X + Y$ tiene distribución Poisson($\lambda_1 + \lambda_2$).

552. Sea X con distribución bin neg(r, p). Demuestre que

a) $G(t) = [p/(1 - t(1 - p))]^r$.

b) $E(X) = r(1 - p)/p$, usando $G(t)$.

c) $\text{Var}(X) = r(1 - p)/p^2$, usando $G(t)$.

Función generadora de momentos

553. Encuentre la función generadora de momentos, si existe, de una variable aleatoria con función de densidad

a) $f(x) = \frac{1}{x!(e - 1)}$, para $x = 1, 2, \dots$

b) $f(x) = e^{-|x|}/2$, para $-\infty < x < \infty$.

554. Sea X con varianza finita y con f.g.m. $M(t)$. Demuestre que

a) $E(X) = M'(0)$.

b) $E(X^2) = M''(0)$.

c) $\text{Var}(X) = M''(0) - (M'(0))^2$.

555. Sean X y Y independientes e idénticamente distribuidas con f.g.m. $M(t)$. Demuestre que $M_{X-Y}(t) = M(t)M(-t)$.

556. Sea X con f.g.m. $M_X(t)$, y sean a y b dos constantes. Demuestre que $M_{aX+b}(t) = e^{tb}M_X(at)$.

557. Sea X con f.g.m. $M_X(t)$. Diga falso o verdadero, demuestre en cada caso.

- a) $M_X(t) \geq 0$.
- b) $M_{2X}(t) = M_X(2t)$.
- c) $M_{X^2}(t) = M_X(tX)$.

558. Sea X con distribución $\text{Ber}(p)$. Demuestre que

- a) $M(t) = 1 - p + pe^t$.
- b) $E(X) = p$, usando $M(t)$.
- c) $E(X^n) = p$, usando $M(t)$.
- d) $\text{Var}(X) = p(1 - p)$, usando $M(t)$.

559. Sea X con distribución $\text{bin}(n, p)$. Demuestre que

- a) $M(t) = (1 - p + pe^t)^n$.
- b) $E(X) = np$, usando $M(t)$.
- c) $\text{Var}(X) = np(1 - p)$, usando $M(t)$.

560. Sean X_1, \dots, X_n independientes cada una con distribución $\text{Ber}(p)$. Use la f.g.m. para demostrar que la variable $X_1 + \dots + X_n$ tiene distribución $\text{bin}(n, p)$.

561. Sean X y Y independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\text{bin}(n + m, p)$.

562. Sea X con distribución $\text{geo}(p)$. Demuestre que

- a) $M(t) = p/[1 - (1 - p)e^t]$.
- b) $E(X) = (1 - p)/p$, usando $M(t)$.
- c) $\text{Var}(X) = (1 - p)/p^2$, usando $M(t)$.

563. Sea X con distribución $\text{Poisson}(\lambda)$. Demuestre que

- a) $M(t) = \exp[\lambda(e^t - 1)]$.
- b) $M''(t) = M'(t) + \lambda e^t M'(t)$.
- c) $E(X) = \lambda$, usando $M(t)$.
- d) $\text{Var}(X) = \lambda$, usando $M(t)$.
- e) $E[(X - \lambda)^3] = \lambda$, usando $M(t)$.
564. Sea X con distribución $\text{unif}(a, b)$. Demuestre que
- a) $M(t) = \frac{e^{bt} - e^{at}}{(b - a)t}$.
- b) $E(X) = (a + b)/2$, usando $M(t)$.
- c) $\text{Var}(X) = (b - a)^2/12$, usando $M(t)$.
565. Sea X con distribución $\exp(\lambda)$. Demuestre que
- a) $M(t) = \lambda/(\lambda - t)$, para $t < \lambda$.
- b) $E(X) = 1/\lambda$, usando $M(t)$.
- c) $\text{Var}(X) = 1/\lambda^2$, usando $M(t)$.
566. Sea X con distribución $N(\mu, \sigma^2)$. Demuestre que
- a) $M(t) = \exp(\mu t + \sigma^2 t^2/2)$.
- b) $E(X) = \mu$, usando $M(t)$.
- c) $\text{Var}(X) = \sigma^2$, usando $M(t)$.
567. Sean X y Y independientes con distribución $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución normal con media $\mu_1 + \mu_2$ y varianza $\sigma_1^2 + \sigma_2^2$.
568. Sea X con distribución $\text{gama}(n, \lambda)$. Demuestre que
- a) $M(t) = [\lambda/(\lambda - t)]^n$, para $t < \lambda$.
- b) $E(X) = n/\lambda$, usando $M(t)$.
- c) $\text{Var}(X) = n/\lambda^2$, usando $M(t)$.

569. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\text{gama}(2, \lambda)$.
570. Sean X y Y independientes con distribución $\text{gama}(n, \lambda)$ y $\text{gama}(m, \lambda)$ respectivamente. Use la f.g.m. para demostrar que la variable $X + Y$ tiene distribución $\text{gama}(n + m, \lambda)$.
571. Sea X con distribución $\chi^2(n)$. Demuestre que
- $M(t) = [1/(1 - 2t)]^{n/2}$, para $t < 1/2$.
 - $E(X) = n$, usando $M(t)$.
 - $\text{Var}(X) = 2n$, usando $M(t)$.
572. Use la f.g.m. para demostrar que si X y Y son independientes tales que X tiene distribución $\chi^2(n)$ y $X + Y$ tiene distribución $\chi^2(m)$ con $m > n$, entonces Y tiene distribución $\chi^2(m - n)$.
573. Sean X y Y independientes con distribución $\chi^2(n)$ y $\chi^2(m)$ respectivamente. Use la f.g.m. para demostrar que $X + Y$ tiene distribución $\chi^2(n + m)$.
574. Sea X con distribución $N(\mu, \sigma^2)$. Use la f.g.m. para demostrar que
- $-X$ tiene distribución $N(-\mu, \sigma^2)$.
 - $aX + b$ tiene distribución $N(a\mu + b, a^2\sigma^2)$, con $a \neq 0$.
 - $(X - \mu)^2/\sigma^2$ tiene distribución $\chi^2(1)$.
575. Sean X_1, \dots, X_n independientes tales que X_k tiene f.g.m. $M_k(t)$ para $k = 1, \dots, n$. Demuestre que $M_{X_1 + \dots + X_n}(t) = M_1(t) \cdots M_n(t)$.
576. Sea X con distribución Cauchy estándar. Demuestre que

$$M_X(t) = \begin{cases} 1 & \text{si } t = 0, \\ \infty & \text{si } t \neq 0. \end{cases}$$

577. Sea X con distribución $t(n)$. Demuestre que

$$M_X(t) = \begin{cases} 1 & \text{si } t = 0, \\ \infty & \text{si } t \neq 0. \end{cases}$$

578. Sea n un número natural. Demuestre que no existe la f.g.m. de la siguiente función de densidad. Esta distribución tiene momentos finitos de orden $1, 2, \dots, n-1$, pero el n -ésimo momento y superiores no existen.

$$f(x) = \begin{cases} n/x^{n+1} & \text{si } x > 1, \\ 0 & \text{otro caso.} \end{cases}$$

Función característica

579. Encuentre la función característica de una variable aleatoria con función de densidad

a) $f(x) = \frac{1}{x!(e-1)}$, para $x = 1, 2, \dots$

b) $f(x) = e^{-|x|}/2$, para $-\infty < x < \infty$.

580. Sea X con función característica $\phi_X(t)$, y sean a y b dos constantes. Demuestre que $\phi_{aX+b}(t) = e^{itb} \phi_X(at)$.
581. Demuestre que una función de distribución $F(x)$ es simétrica si, y sólo si, la correspondiente función característica $\phi(t)$ es real.
582. Demuestre que la función característica es una función uniformemente continua, es decir, para todo $\epsilon > 0$ existe $\delta > 0$ tal que para todo t y s con $|t - s| < \delta$, se cumple que $|\phi(t) - \phi(s)| < \epsilon$.
583. Demuestre que la función característica satisface la igualdad $\phi(-t) = \overline{\phi(t)}$, en donde \overline{z} denota el complejo conjugado de z .
584. Sean $\phi_1(t)$ y $\phi_2(t)$ dos funciones características, y sea $\alpha \in [0, 1]$. Demuestre que la combinación lineal convexa $\alpha\phi_1(t) + (1 - \alpha)\phi_2(t)$ es una función característica.
585. Sean X y Y independientes y con idéntica distribución. Demuestre que $\phi_{X-Y}(t) = |\phi_X(t)|^2$, en este caso la función característica es una función real por que la variable $X - Y$ es simétrica.

586. Sea X con distribución $\text{Ber}(p)$. Demuestre que

- a) $\phi(t) = 1 - p + pe^{it}$.
- b) $E(X) = p$, usando $\phi(t)$.
- c) $\text{Var}(X) = p(1 - p)$, usando $\phi(t)$.
- d) $E(X^n) = p$, usando $\phi(t)$, con $n \geq 1$ entero.

587. Sea X con distribución $\text{bin}(n, p)$. Hemos demostrado que la función característica de esta distribución es $\phi(t) = (1 - p + pe^{it})^n$. Usando $\phi(t)$ demuestre ahora que

- a) $E(X) = np$.
- b) $E(X^2) = np(1 - p + np)$.
- c) $\text{Var}(X) = np(1 - p)$.

588. Sea X con distribución $\text{Poisson}(\lambda)$. Hemos demostrado que la función característica de esta distribución es $\phi(t) = \exp[-\lambda(1 - e^{it})]$. Usando $\phi(t)$ compruebe que

- a) $E(X) = \lambda$.
- b) $E(X^2) = \lambda(\lambda + 1)$.
- c) $\text{Var}(X) = \lambda$.

589. Sea X con distribución $\text{geo}(p)$. Demuestre que

- a) $\phi(t) = p/(1 - (1 - p)e^{it})$.
- b) $E(X) = (1 - p)/p$, usando $\phi(t)$.
- c) $\text{Var}(X) = (1 - p)/p^2$, usando $\phi(t)$.

590. Sea X tiene distribución $\text{bin neg}(r, p)$. Demuestre que

- a) $\phi(t) = [p/(1 - (1 - p)e^{it})]^r$.
- b) $E(X) = r(1 - p)/p$, usando $\phi(t)$.
- c) $\text{Var}(X) = r(1 - p)/p^2$, usando $\phi(t)$.

591. Sea X con distribución $\text{unif}(-a, a)$. Demuestre que $\phi(t) = (\text{sen } at)/at$.

592. Sea X con distribución $\text{unif}(a, b)$. Demuestre que

a) $\phi(t) = [e^{ibt} - e^{iat}]/[it(b - a)]$.

b) $E(X) = (a + b)/2$, usando $\phi(t)$.

c) $\text{Var}(X) = (b - a)^2/12$, usando $\phi(t)$.

593. Sea X con distribución $N(\mu, \sigma^2)$. Hemos demostrado que la función característica de esta distribución es $\phi(t) = \exp(i\mu t - \sigma^2 t^2/2)$. Usando $\phi(t)$ compruebe que $E(X) = \mu$ y $\text{Var}(X) = \sigma^2$.

594. Sea X con distribución normal estándar. Use la función característica para demostrar que para $n = 0, 1, \dots$

$$E(X^n) = \begin{cases} \frac{n!}{2^{n/2}(n/2)!} & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

595. Sea X con distribución $\exp(\lambda)$. Demuestre que $\phi(t) = \lambda/(\lambda - it)$. Use $\phi(t)$ para comprobar que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$.

596. Sea X con distribución $\text{gama}(n, \lambda)$. Hemos encontrado que la función característica de esta distribución es $\phi(t) = [\lambda/(\lambda - it)]^n$. Usando $\phi(t)$ compruebe nuevamente que

a) $E(X) = n/\lambda$.

b) $E(X^m) = \frac{\Gamma(m + n)}{\lambda^m \Gamma(n)}$, para $m = 0, 1, \dots$

c) $\text{Var}(X) = n/\lambda^2$.

597. Sean X y Y independientes ambas con distribución $\exp(\lambda)$. Use la función característica para demostrar que la variable $X + Y$ tiene distribución $\text{gama}(2, \lambda)$.

598. Sean X y Y independientes con distribución $\text{gama}(n, \lambda)$ y $\text{gama}(m, \lambda)$ respectivamente. Use la función característica para demostrar que la variable $X + Y$ tiene distribución $\text{gama}(n + m, \lambda)$.

599. Sea X con función de distribución $F(x) = e^x/(1 + e^x)$. Demuestre que $F(x)$ es efectivamente una función de distribución, y calcule su función característica asociada. Con ayuda de ésta última encuentre la esperanza y la varianza de X .
600. Sean X y Y independientes. Demuestre que

$$\phi_{XY}(t) = \int_{-\infty}^{\infty} \phi_Y(tx) dF_X(x) = \int_{-\infty}^{\infty} \phi_X(ty) dF_Y(y).$$

601. Mediante el *cálculo de residuos* de la teoría de variable compleja puede demostrarse que la distribución Cauchy estándar tiene función característica

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\pi(1+x^2)} dx = e^{-|t|}.$$

Suponiendo este resultado, encuentre el error en el siguiente argumento para encontrar la f.g.m. de la distribución Cauchy: “Como $\phi(t) = e^{-|t|}$ y $M(t) = \phi(-it)$, entonces $M(t) = e^{-|-it|} = e^{-|t|}$.” El caso es que no existe la f.g.m. para la distribución Cauchy.

602. Sean X_1, \dots, X_n independientes cada una de ellas con distribución Cauchy estándar, es decir, la función característica es $\phi(t) = e^{-|t|}$. Use este resultado para demostrar que la v.a. $S_n = (X_1 + \dots + X_n)/n$ tiene distribución Cauchy estándar para cualquier valor de n .

CAPÍTULO 9

Dos teoremas límite

En este último capítulo se estudian dos de los teoremas más importantes en probabilidad: la ley de los grandes números y el teorema central del límite. Antes de ello se revisan algunas desigualdades de interés general.

9.1. Algunas desigualdades

PROPOSICIÓN. (DESIGUALDAD DE MARKOV). Sea $X \geq 0$ una variable aleatoria con esperanza finita. Para cualquier $\epsilon > 0$,

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}.$$

Demostración.

$$\begin{aligned}
 E(X) &= E(X 1_{(X \geq \epsilon)} + X 1_{(X < \epsilon)}) \\
 &\geq E(X 1_{(X \geq \epsilon)}) \\
 &\geq E(\epsilon 1_{(X \geq \epsilon)}) \\
 &= \epsilon P(X \geq \epsilon).
 \end{aligned}$$

□

En palabras, este resultado establece que la probabilidad de que X exceda un valor ϵ positivo está acotada superiormente por la media entre ϵ . Existen otras versiones equivalentes de esta desigualdad, por ejemplo,

- a) $P(|X| \geq \epsilon) \leq E|X|/\epsilon$.
- b) $P(|X| \geq \epsilon) \leq E|X|^n/\epsilon^n$, con n en \mathbb{N} .

La siguiente desigualdad será usada en la siguiente sección para demostrar la ley débil de los grandes números.

PROPOSICIÓN. (DESIGUALDAD DE CHEBYSHEV). Sea X una variable aleatoria con media μ y varianza finita σ^2 . Para cualquier $\epsilon > 0$,

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \quad (9.1)$$

Demostración.

$$\begin{aligned}
 \sigma^2 &= E[(X - \mu)^2] \\
 &= E[(X - \mu)^2 1_{(|X - \mu| \geq \epsilon)} + (X - \mu)^2 1_{(|X - \mu| < \epsilon)}] \\
 &\geq E[(X - \mu)^2 1_{(|X - \mu| \geq \epsilon)}] \\
 &\geq E[\epsilon^2 1_{(|X - \mu| \geq \epsilon)}] \\
 &= \epsilon^2 P(|X - \mu| \geq \epsilon).
 \end{aligned}$$

□

En palabras, esta desigualdad dice que la probabilidad de que X difiera de su media en más de ϵ está acotada superiormente por la varianza entre ϵ^2 . A este resultado se le conoce también con el nombre de *desigualdad de Chebyshev-Bienaymé*. Existen otras versiones de esta desigualdad equivalentes a la demostrada, por ejemplo,

- a) $P(|X - \mu| \geq \epsilon\sigma) \leq 1/\epsilon^2$.
- b) $P(|X - \mu| < \epsilon\sigma) \geq 1 - 1/\epsilon^2$.
- c) $P(|X - \mu| < \epsilon) \geq 1 - \sigma^2/\epsilon^2$.

Ahora demostraremos una versión de la desigualdad de Chebyshev un poco más general.

PROPOSICIÓN. (DESIGUALDAD DE CHEBYSHEV EXTENDIDA). Sea X una variable aleatoria, y sea $g \geq 0$ una función no decreciente tal que $g(X)$ es una variable aleatoria con esperanza finita. Para cualquier $\epsilon > 0$,

$$P(X \geq \epsilon) \leq \frac{E[g(X)]}{g(\epsilon)}. \quad (9.2)$$

Demostración.

$$\begin{aligned} E[g(X)] &= E[g(X) 1_{(X \geq \epsilon)} + g(X) 1_{(X < \epsilon)}] \\ &\geq E[g(X) 1_{(X \geq \epsilon)}] \\ &\geq E[g(\epsilon) 1_{(X \geq \epsilon)}] \\ &= g(\epsilon)P(X \geq \epsilon). \end{aligned}$$

□



Pafnuty Lvovich Chebyshev
(Rusia, 1821–1894)



Andrei Andreyevich Markov
(Rusia, 1856–1922)

Profesor y alumno.

Fuente: Archivo MacTutor, Universidad de St. Andrews.

A partir de la desigualdad anterior y con una función g adecuada se pueden obtener tanto la desigualdad de Chebyshev como la desigualdad de Markov.

PROPOSICIÓN. (DESIGUALDAD DE KOLMOGOROV). Sean X_1, \dots, X_n independientes con media cero y segundo momento finito. Para cualquier $\epsilon > 0$,

$$P(\max_k \{|X_1 + \dots + X_k|\} \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{k=1}^n \text{Var}(X_k).$$

Demostración. Para cada $k = 1, \dots, n$, defina $S_k = X_1 + \dots + X_k$, cuya esperanza es cero por hipótesis. Observe que las variables S_k y $S_n - S_k$ son independientes y por lo tanto $E(S_k(S_n - S_k)) = 0$. Defina ahora los eventos disjuntos

$$A_k = (|S_k| \geq \epsilon) \cap \bigcap_{i=1}^{k-1} (|S_i| < \epsilon),$$

en donde en particular $A_1 = (|S_1| \geq \epsilon)$. El evento de interés puede escribirse como $A = \bigcup_{k=1}^n A_k$. Entonces

$$\begin{aligned}
 E(S_n^2) &\geq E(S_n^2 1_A) = \sum_{k=1}^n E(S_n^2 1_{A_k}) \\
 &= \sum_{k=1}^n E((S_k + (S_n - S_k))^2 1_{A_k}) \\
 &= \sum_{k=1}^n E(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2 1_{A_k}) \\
 &\geq \sum_{k=1}^n E(S_k^2 1_{A_k}) \geq \sum_{k=1}^n \epsilon^2 E(1_{A_k}) \geq \sum_{k=1}^n \epsilon^2 P(A_k) \\
 &= \epsilon^2 P(A).
 \end{aligned}$$

El resultado se obtiene al observar que $E(S_n^2) = \text{Var}(S_n) = \sum_{k=1}^n \text{Var}(X_k)$. \square

Cuando $n = 1$ la desigualdad de Kolmogorov se reduce a la desigualdad de Chebyshev. En resumen se tiene la siguiente tabla.

ALGUNAS DESIGUALDADES	
MARKOV:	a) $P(X \geq \epsilon) \leq E(X)/\epsilon$, para $X \geq 0$. b) $P(X \geq \epsilon) \leq E X /\epsilon$. c) $P(X \geq \epsilon) \leq E X ^n/\epsilon^n$.
CHEBYSHEV:	a) $P(X - \mu \geq \epsilon) \leq \text{Var}(X)/\epsilon^2$. b) $P(X \geq \epsilon) \leq E[g(X)]/g(\epsilon)$, con $g \geq 0$ no decreciente.
KOLMOGOROV:	$P(\max_k \{ X_1 + \dots + X_k \} \geq \epsilon) \leq \frac{1}{\epsilon^2} \sum_{k=1}^n \text{Var}(X_k)$.

9.2. Ley de los grandes números

Este interesante resultado establece que, bajo ciertas condiciones, el promedio de variables aleatorias converge a una constante cuando el número de sumandos crece a infinito. Demostraremos dos versiones de esta afirmación, las cuales se distinguen por el tipo de convergencia de la que se trate. La *ley débil* establece la convergencia en probabilidad y la *ley fuerte* dice que la convergencia es casi segura. La ley fuerte implica entonces la ley débil. Existen además varias generalizaciones de este resultado.

TEOREMA DE BERNOULLI. (LEY DÉBIL DE LOS GRANDES NÚMEROS). Sean X_1, X_2, \dots independientes e idénticamente distribuidas con media μ . Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Demostración. Sea $S_n = (X_1 + \dots + X_n)/n$, y sea $\phi(t)$ la función característica de cualquier elemento X de la sucesión. Como X tiene esperanza finita μ y por la expansión (8.1),

$$\phi(t) = 1 + it(\mu + o(1)), \quad \text{cuando } t \rightarrow 0.$$

Por independencia la función característica de S_n es entonces

$$\phi_{S_n}(t) = \phi^n(t/n) = (1 + i(t/n)(\mu + o(1)))^n, \quad \text{cuando } t \rightarrow 0,$$

Haciendo $n \rightarrow \infty$ se obtiene $\phi_{S_n}(t) \rightarrow e^{i\mu t}$, en donde $e^{i\mu t}$ es la función característica de la variable aleatoria constante μ . Esto implica que $S_n \xrightarrow{d} \mu$. El resultado se obtiene al recordar que la convergencia en distribución a una constante es equivalente a la convergencia en probabilidad. \square

Este mismo resultado puede demostrarse fácilmente a partir de la desigualdad de Chebyshev bajo la hipótesis adicional de existencia de la varianza.

El argumento es el siguiente. Sea nuevamente $S_n = (X_1 + \cdots + X_n)/n$. Entonces $E(S_n) = \mu$ y $\text{Var}(S_n) = \sigma^2/n$, suponiendo $\text{Var}(X) = \sigma^2 < \infty$. La desigualdad de Chebyshev aplicada a la variable S_n asegura que para cualquier $\epsilon > 0$ se cumple $P(|S_n - \mu| \geq \epsilon) \leq \sigma^2/n\epsilon^2$. Basta ahora tomar el límite cuando n tiende a infinito para obtener el resultado.

Damos a continuación un ejemplo sencillo de aplicación de la ley débil y más adelante demostramos la ley fuerte.

EJEMPLO (PROBABILIDAD FRECUENTISTA). Considere un experimento aleatorio cualquiera y sea A un evento. Se efectúan realizaciones independientes del experimento, y se observa en cada ensayo la ocurrencia o no ocurrencia del evento A . Sea X_k la variable que toma el valor uno si en el k -ésimo ensayo se observa A , y cero en caso contrario. Entonces las variables X_1, X_2, \dots son independientes cada una con distribución $\text{Ber}(p)$, en donde p es la probabilidad desconocida del evento A . Por lo tanto $E(X_k) = p$ y $\text{Var}(X_k) = p(1-p)$. La ley débil de los grandes números asegura que la fracción de ensayos en los que se observa el evento A converge, en probabilidad, a la constante desconocida p cuando el número de ensayos crece a infinito. Esta es la definición frecuentista de la probabilidad, y hemos entonces corroborado su validez con ayuda de la ley de los grandes números. \bullet

EJEMPLO. A continuación se muestra gráficamente una simulación en computadora del comportamiento del cociente $(X_1 + \cdots + X_n)/n$ cuando n crece. Se muestra también el código MATLAB utilizado, el cual puede ser traducido fácilmente a cualquier otro lenguaje de programación. Se generaron 200 valores al azar usando la distribución discreta $\text{Ber}(p)$, con $p = 0.5$. El comando “binornd(n,p)” genera un valor al azar de la distribución $\text{bin}(n,p)$. Los datos obtenidos por este paquete fueron luego trasladados a L^AT_EX, usando pstricks, para generar la gráfica mostrada en la Figura 9.1. Los puntos graficados fueron unidos por una línea continua para una mejor visualización del comportamiento inicial oscilante y su eventual estabilización. \bullet

EJEMPLO. Esta es otra simulación en computadora del comportamiento del cociente $(X_1 + \cdots + X_n)/n$ cuando n crece, ahora usando la distribución

```

randn('state',150)
N=200; S=zeros(1,N); Sn=zeros(1,N);
p=0.5; R=binornd(1,p);
S(1)=R; Sn(1)=R;
for j=2:N
    S(j)=S(j-1)+binornd(1,p);
    Sn(j)=S(j)/j;
end
plot([Sn],'r-')

```

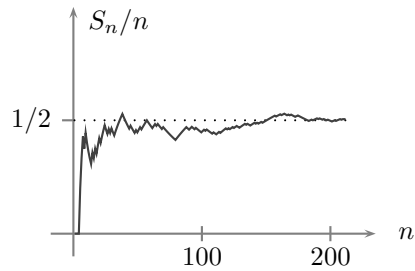


Figura 9.1: Comportamiento del cociente S_n/n cuando n crece cuando las variables X_i tienen distribución discreta $\text{Ber}(p)$, con $p = 0.5$, y el código MATLAB para generar la simulación.

continua $N(1, 9)$. El comando “randn” genera valores al azar de la distribución normal estándar, de modo que la expresión “ $1+3*\text{randn}$ ” corresponde a un valor de la distribución $N(1, 9)$. Se generaron nuevamente 200 de estos valores y los resultados de muestran en la Figura 9.2. Es gratificante observar las oscilaciones iniciales de dicho cociente y su eventual estabilización hacia la media de la distribución. ■

TEOREMA. (LEY FUERTE DE LOS GRANDES NÚMEROS). Sean X_1, X_2, \dots independientes e idénticamente distribuidas con media μ . Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{c.s.}} \mu.$$

Demostración. (Suponiendo cuarto momento finito). Dada la idéntica distribución de los elementos de la sucesión, cualquier elemento de ésta se denota simplemente por X . Suponga que $E|X - \mu|^2 = \sigma^2$ y observe que

```

randn('state',1500)
N=200; S=zeros(1,N); Sn=zeros(1,N);
R=1+3*randn;
S(1)=R; Sn(1)=R;
for j=2:N
    S(j)=S(j-1)+1+3*randn;
    Sn(j)=S(j)/j;
end
plot([Sn], 'r-')

```

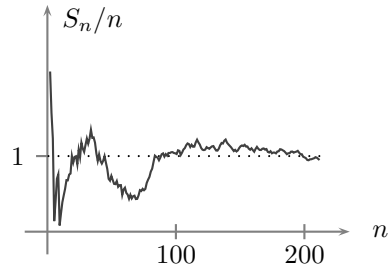


Figura 9.2: Comportamiento del cociente S_n/n cuando n crece usando la distribución normal con media uno y varianza nueve.

$E(X - \mu) = 0$. Entonces por independencia,

$$E\left|\sum_{i=1}^n (X_i - \mu)\right|^4 = nE|X - \mu|^4 + 3n(n-1)\sigma^4.$$

Por la desigualdad de Chebyshev (9.2) aplicada a la variable $|\sum_{i=1}^n (X_i - \mu)|$ y la función $g(x) = x^4$ se obtiene, para $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{i=1}^n (X_i - \mu)\right| > n\epsilon\right) &\leq E\left|\sum_{i=1}^n (X_i - \mu)\right|^4 / (n\epsilon)^4 \\ &= (nE|X - \mu|^4 + 3n(n-1)\sigma^4) / (n\epsilon)^4. \end{aligned}$$

Sea el evento $A_n = (|\frac{1}{n} \sum_{i=1}^n X_i - \mu| > \epsilon)$. Entonces $\sum_{n=1}^{\infty} P(A_n) < \infty$. Por el lema de Borel-Cantelli la probabilidad de que ocurra una infinidad de eventos A_n es cero, es decir, con probabilidad uno, sólo un número finito de estos eventos ocurre. Por lo tanto con probabilidad uno, existe un número natural n a partir del cual ningún evento A_n se verifica. Es decir,

$$P\left(\lim_{n \rightarrow \infty} \left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \leq \epsilon\right) = 1.$$

Como esta afirmación vale para cualquier $\epsilon > 0$, se cumple que

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu\right) = 1.$$

□

EJEMPLO. (EL PROBLEMA DEL MONO, NUEVAMENTE). Usaremos la ley fuerte de los grandes números para dar otra solución al problema del mono. Considere entonces un mono que escribe caracteres al azar. Nos interesa encontrar la probabilidad de que el mono eventualmente escriba las obras completas de Shakespeare, las cuales, supondremos, tienen una longitud total de N caracteres. Nuevamente se consideran bloques de longitud N de la siguiente forma

$$\underbrace{x_1, \dots, x_N}_{\text{bloque 1}}, \underbrace{x_{N+1}, \dots, x_{2N}}_{\text{bloque 2}}, \dots$$

Sea A_k el evento correspondiente a que en el k -ésimo bloque el mono tenga éxito, y sea X_k la variable aleatoria indicadora del evento A_k , es decir,

$$X_k = \begin{cases} 1 & \text{si } A_k \text{ ocurre,} \\ 0 & \text{si } A_k \text{ no ocurre.} \end{cases}$$

Se tiene entonces una sucesión de variables aleatorias X_1, X_2, \dots independientes e idénticamente distribuidas $\text{Ber}(p)$, con $p = P(A_k) = (1/m)^N$, suponiendo que el total de caracteres disponibles es m . En particular, la media de cada una de estas variables es $E(X_k) = p$. Considere ahora la suma $X_1 + \dots + X_n$. Si para algún valor de n esta suma es positiva, significa que alguno de los sumandos es distinto de cero, y por lo tanto que el mono ha tenido éxito. Pero esto es justamente lo que garantiza la ley fuerte de los grandes números pues

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = p\right) = 1.$$

Es decir, con probabilidad uno la suma en esta ecuación es positiva. Esto implica que debe existir un valor de k tal que $X_k = 1$, y esto a su vez

significa que en el k -ésimo bloque el mono ha tenido éxito. Más aún, para que el promedio que aparece en esta ecuación sea positivo necesariamente la suma debe ser infinita, y por lo tanto, deben existir una infinidad de valores de k tal que $X_k = 1$. Esto quiere decir que con probabilidad uno el mono escribirá una infinidad de veces las obras completas de Shakespeare. \square

9.3. Teorema central del límite

Concluimos el curso con el célebre y famoso teorema central del límite. Este resultado es de amplio uso en estadística y otras ramas de aplicación de la probabilidad. Existen muchas versiones y generalizaciones de este teorema pero nos limitaremos a enunciar y demostrar una versión simple y corta. Un caso particular de este resultado lleva el nombre de A. de Moivre y de P. S. Laplace.

TEOREMA DE DE MOIVRE-LAPLACE. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes tal que cada una de ellas tiene distribución Bernoulli con parámetro $p \in (0, 1)$. Para cualesquiera números reales $a < b$,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

En palabras este resultado establece que la variable aleatoria $(X_1 + \dots + X_n - np)/\sqrt{np(1-p)}$ converge en distribución a una variable aleatoria normal estándar, una demostración directa puede ser encontrada en [8]. Este teorema fue descubierto por A. de Moivre alrededor de 1733 en el caso cuando las variables aleatorias tienen distribución Bernoulli con $p = 1/2$. Años después P. S. Laplace demostró su validez para valores arbitrarios de p . El teorema de de Moivre-Laplace es un caso particular del siguiente resultado fundamental.

TEOREMA CENTRAL DEL LÍMITE. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas tales que para cada natural n , $E(X_n) = \mu$ y $\text{Var}(X_n) = \sigma^2 < \infty$. Entonces

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1).$$

Demostración. Observe que

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{(X_1 - \mu)/\sigma + \dots + (X_n - \mu)/\sigma}{\sqrt{n}},$$

en donde cada sumando del numerador en el lado derecho es una variable con media cero y varianza uno. Así pues, sin pérdida de generalidad, supondremos que cada variable de la sucesión tiene media cero y varianza uno. Considere entonces la suma $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$. Se desea probar que $Z_n \xrightarrow{d} N(0, 1)$. Para ello es suficiente demostrar que $\phi_{Z_n}(t) \rightarrow e^{-t^2/2}$. Por independencia e idéntica distribución,

$$\phi_{Z_n}(t) = E(e^{it(X_1 + \dots + X_n)/\sqrt{n}}) = (\phi_X(t/\sqrt{n}))^n,$$

en donde $\phi_X(t)$ es la función característica de cualquier elemento de la sucesión, que por la expansión (8.1) adquiere la expresión, cuando $t \rightarrow 0$,

$$\phi_X(t) = 1 - \frac{1}{2}t^2(1 + o(1)).$$

Por lo tanto,

$$\phi_{Z_n}(t) = \left(1 - \frac{t^2}{2n}(1 + o(1))\right)^n.$$

Haciendo $n \rightarrow \infty$ se obtiene $\phi_{Z_n}(t) \rightarrow e^{-t^2/2}$. □

El teorema central del límite establece entonces que para cualquier número real x ,

$$\lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = P(Z \leq x),$$

en donde Z tiene distribución normal estándar. Observe que la suma $X_1 + \cdots + X_n$ tiene media $n\mu$ y varianza $n\sigma^2$, de modo que la expresión de arriba es una especie de estandarización de esta variable. Equivalentemente el resultado puede enunciarse del siguiente modo:

$$\frac{(X_1 + \cdots + X_n)/n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1).$$

Este teorema fue demostrado rigurosamente por A. M. Lyapunov alrededor de 1901. Observe que no hay ninguna hipótesis adicional sobre la distribución de las variables de la sucesión, es decir, éstas pueden tener cualquier distribución, sólo requiriendo la existencia de la media y la varianza.

9.4. Ejercicios

Desigualdad de Markov

603. Demuestre la desigualdad de Markov siguiendo los siguientes pasos: Suponga $X \geq 0$, y para $\epsilon > 0$ defina

$$X_\epsilon = \begin{cases} \epsilon & \text{si } X \geq \epsilon, \\ 0 & \text{si } X < \epsilon. \end{cases}$$

Compruebe que $X_\epsilon \leq X$. Ahora tome esperanza de ambos lados y calcule $E(X_\epsilon)$.

604. Use la desigualdad de Markov para demostrar que si X es una variable aleatoria no negativa con esperanza cero, entonces $X = 0$ casi seguramente.
605. CONV. EN MEDIA \Rightarrow CONV. EN PROBABILIDAD. Demuestre que la convergencia en media implica la convergencia en probabilidad, usando la desigualdad de Markov aplicada a la variable aleatoria no negativa $|X_n - X|$.

Desigualdad de Chebyshev

606. CONV. EN M.C. \Rightarrow CONV. EN PROBABILIDAD. Use la desigualdad de Chebyshev (9.2) para demostrar directamente que la convergencia en media cuadrática implica la convergencia en probabilidad.
607. Demuestre la desigualdad de Chebyshev (9.1) usando la desigualdad de Markov aplicada a la variable aleatoria no negativa $|X - \mu|$.
608. Use la desigualdad de Chebyshev para demostrar que si X es una variable aleatoria tal que $E(X) = a$ y $\text{Var}(X) = 0$, entonces X es constante casi seguramente, es decir, $P(X = a) = 1$.

609. Sea X con media μ y varianza σ^2 . Use la desigualdad de Chebyshev para estimar la probabilidad de que X tome valores entre $\mu - \epsilon\sigma$ y $\mu + \epsilon\sigma$ para cualquier $\epsilon > 0$ constante.
610. A partir de la desigualdad de Chebyshev extendida (9.2) demuestre la desigualdad de Chebyshev (9.1) y la desigualdad de Markov.
611. Demuestre que $P(|X| \geq \epsilon) \leq E|X|/\epsilon$, para $\epsilon > 0$,
- usando la desigualdad de Chebyshev extendida.
 - de manera directa.
612. Demuestre que $P(|X| \geq \epsilon) \leq E|X|^n/\epsilon^n$, para $\epsilon > 0$ y $n \in \mathbb{N}$,
- usando la desigualdad de Chebyshev extendida.
 - de manera directa.
613. Demuestre que $P(X \geq \epsilon) \leq E(e^{tX})/e^{\epsilon t}$, para $\epsilon > 0$ y $t > 0$,
- usando la desigualdad de Chebyshev extendida.
 - de manera directa.
614. Sea X discreta con función de probabilidad

$$f(x) = \begin{cases} 1/18 & \text{si } x = -1, 1, \\ 16/18 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases}$$

Demuestre que el valor exacto de la probabilidad $P(|X - \mu| \geq 3\sigma)$ coincide con la estimación dada por la desigualdad de Chebyshev. Este resultado demuestra que, sin hipótesis adicionales, la cota superior dada por la desigualdad de Chebyshev es óptima.

615. Considere la siguiente versión de la desigualdad de Chebyshev

$$P(|X - \mu| < \epsilon\sigma) \geq 1 - 1/\epsilon^2.$$

Encuentre el mínimo valor de $\epsilon > 0$ de tal modo que la probabilidad de que una variable aleatoria tome valores entre $\mu - \epsilon\sigma$ y $\mu + \epsilon\sigma$ sea al menos 0.90.

616. DESIGUALDAD DE CANTELLI. Demuestre que si $\text{Var}(X) < \infty$, entonces para cualquier $\epsilon > 0$,

$$P(|X - E(X)| > \epsilon) \leq \frac{2 \text{Var}(X)}{\epsilon^2 + \text{Var}(X)}.$$

Ley de los grandes números

617. Use la ley débil de los grandes números para demostrar que si X_n tiene distribución $\text{bin}(n, p)$, entonces $\frac{1}{n}X_n \xrightarrow{p} p$, cuando n tiende a infinito.
618. LEY DE LOS GRANDES NÚMEROS EN MEDIA CUADRÁTICA. Demuestre que si X_1, X_2, \dots son independientes con media μ y varianza σ^2 , entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{m.c.} \mu.$$

Observe que no se pide la hipótesis de idéntica distribución para las variables aleatorias y que este resultado no es consecuencia de la ley fuerte.

619. Sean X_1, \dots, X_n independientes con distribución $N(\mu, \sigma^2)$. El promedio $(X_1 + \dots + X_n)/n$ tiene distribución $N(\mu, \sigma^2/n)$ para cualquier valor de n . ¿Contradice esto la ley de los grandes números?
620. En el ejercicio 602 se pide usar la función característica para demostrar que si X_1, \dots, X_n son independientes con distribución Cauchy estándar, entonces el promedio $S_n = (X_1 + \dots + X_n)/n$ tiene distribución Cauchy estándar, independientemente del valor de n . ¿Contradice esto la ley de los grandes números?
621. Se lanza una moneda equilibrada $2n$ veces. Calcule la probabilidad de que ambas caras caigan el mismo número de veces. ¿Qué le sucede a esta probabilidad cuando n tiende a infinito? ¿Contradice esto la ley de los grandes números?

Teorema central del límite

622. Use el teorema central del límite para estimar la probabilidad de obtener mas de 520 águilas en 1000 lanzamientos de una moneda honesta.
623. Sean X_1, X_2, \dots independientes con distribución Poisson(λ) con $\lambda = 1$. Use el teorema central del límite para demostrar que

$$\lim_{n \rightarrow \infty} \frac{1}{e^n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}.$$

624. La probabilidad de ocurrencia de un evento en un ensayo es de 0.3. ¿Cuál es la probabilidad de que la frecuencia relativa de este evento en 100 ensayos se encuentre entre 0.2 y 0.5?

APÉNDICE A

Distribuciones de probabilidad

Se presenta a continuación una lista en orden alfabético de algunas distribuciones de probabilidad univariadas de uso común. Como es costumbre, la función de probabilidad o de densidad se denota por $f(x)$, y la función de distribución por $F(x)$. Como en el texto, $G(t)$ es la función generadora de probabilidad, $M(t)$ es la función generadora de momentos, y $\phi(t)$ es la función característica.

Distribución Bernoulli

$X \sim \text{Ber}(p)$, con $p \in (0, 1)$.

$f(x) = p^x(1-p)^{1-x}$ para $x = 0, 1$.

$E(X) = p$.

$\text{Var}(X) = p(1-p)$.

$G(t) = 1 - p + pt$.

$M(t) = 1 - p + pe^t$.

Este es el modelo más simple de variable aleatoria y corresponde a la observación de la ocurrencia o no ocurrencia de un evento. La suma de n variables independientes $\text{Ber}(p)$ tiene distribución $\text{bin}(n, p)$.

Distribución beta

$X \sim \text{beta}(a, b)$ con $a > 0, b > 0$.

$f(x) = x^{a-1}(1-x)^{b-1}/B(a, b)$, para $x \in (0, 1)$.

$E(X) = a/(a+b)$.

$\text{Var}(X) = ab/[(a+b+1)(a+b)^2]$.

Cuando $a = 1, b = 2$ o $a = 2, b = 1$ se obtiene la distribución triangular.

Distribución binomial

$X \sim \text{bin}(n, p)$ con $n \in \mathbb{N}$ y $p \in (0, 1)$.

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ para $x = 0, 1, \dots, n$.

$E(X) = np$.

$\text{Var}(X) = np(1-p)$.

$G(t) = (1-p+pt)^n$.

$M(t) = [1-p+pe^t]^n$.

Una variable aleatoria binomial registra el número de éxitos en n ensayos independientes Bernoulli en donde en cada ensayo la probabilidad de éxito es p . La suma de dos variables independientes con distribución $\text{bin}(n, p)$ y $\text{bin}(m, p)$ tiene distribución $\text{bin}(n+m, p)$.

Distribución binomial negativa

$X \sim \text{bin neg}(r, p)$ con $r \in \mathbb{N}$ y $p \in (0, 1)$.

$f(x) = \binom{r+x-1}{x} p^r (1-p)^x$ para $x = 0, 1, \dots$

$E(X) = r(1-p)/p$.

$\text{Var}(X) = r(1-p)/p^2$.

$G(t) = [p/(1-t(1-p))]^r$.

$$M(t) = [p/(1 - qe^t)]^r.$$

Este es el modelo que se usa para contar el número de fracasos antes de obtener el r -ésimo éxito en una sucesión de ensayos independientes Bernoulli, en donde en cada ensayo la probabilidad de éxito es p . La distribución binomial negativa se reduce a la distribución geométrica cuando $r = 1$.

Distribución Cauchy

$X \sim \text{Cauchy}(a, b)$ con $a > 0$ y $b > 0$.

$$f(x) = \frac{1}{b\pi[1 + ((x - a)/b)^2]}.$$

La esperanza, la varianza y cualquier momento no existen.

La función generadora de momentos no existe para $t \neq 0$.

$$\phi(t) = \exp(iat - b|t|).$$

Cuando $a = 0$ y $b = 1$ se obtiene la distribución Cauchy estándar, y coincide con la distribución $t(n)$ con $n = 1$. En este caso,

$$f(x) = 1/(\pi(1 + x^2)), \quad \text{para } x \in \mathbb{R}.$$

$$F(x) = 1/2 + (\arctan x)/\pi, \quad \text{para } x \in \mathbb{R}.$$

Distribución exponencial

$X \sim \exp(\lambda)$ con $\lambda > 0$.

$$f(x) = \lambda e^{-\lambda x}, \quad \text{para } x > 0.$$

$$F(x) = 1 - e^{-\lambda x}, \quad \text{para } x > 0.$$

$$E(X) = 1/\lambda.$$

$$\text{Var}(X) = 1/\lambda^2.$$

$$M(t) = \lambda/(\lambda - t) \quad \text{para } t < \lambda.$$

$$\phi(t) = \lambda/(\lambda - it).$$

La suma de n variables independientes $\exp(\lambda)$ tiene distribución $\text{gama}(n, \lambda)$.

Distribución gama

$X \sim \text{gama}(n, \lambda)$ con $n > 0$ y $\lambda > 0$.

$$f(x) = \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x} \quad \text{para } x > 0.$$

$$F(x) = 1 - e^{-\lambda x} \sum_{k=0}^{n-1} (\lambda x)^k / k! \quad \text{para } x > 0 \text{ y } n \text{ entero.}$$

$$E(X) = n/\lambda.$$

$$\text{Var}(X) = n/\lambda^2.$$

$$M(t) = [\lambda/(\lambda - t)]^n, \quad \text{para } t < \lambda.$$

Cuando $n = 1$ la distribución gama se reduce a la distribución exponencial. Advertencia: para denotar esta distribución en algunos textos se usa el símbolo $\text{gama}(\lambda, n)$, es decir, el orden de los parámetros es distinto. En ocasiones se usa el parámetro $1/\theta$ en lugar de λ .

Distribución geométrica

$X \sim \text{geo}(p)$ con $p \in (0, 1)$.

$$f(x) = p(1 - p)^x \quad \text{para } x = 0, 1, \dots$$

$$E(X) = (1 - p)/p.$$

$$\text{Var}(X) = (1 - p)/p^2.$$

$$G(t) = p/[1 - t(1 - p)].$$

$$M(t) = p/[1 - (1 - p)e^t].$$

Esta variable se usa para modelar el número de fracasos antes de obtener el primer éxito en una sucesión de ensayos independientes Bernoulli, en donde en cada uno de ellos la probabilidad de éxito es p . La distribución geométrica es un caso particular de la distribución binomial negativa.

Distribución hipergeométrica

$X \sim \text{hipergeo}(N, K, n)$ con $N, K, n \in \mathbb{N}$ y $n \leq K \leq N$.

$$f(x) = \binom{K}{x} \binom{N-K}{n-x} / \binom{N}{n} \quad \text{para } x = 0, 1, \dots, n.$$

$$E(X) = nK/N.$$

$$\text{Var}(X) = n \frac{K}{N} \frac{N-K}{N} \frac{N-n}{N-1}.$$

Si un conjunto de N elementos se puede separar en dos clases, una clase con K elementos y la otra con $N - K$ elementos, y si se seleccionan n elementos de este conjunto, entonces la variable X modela el número de elementos seleccionados de la primera clase.

Distribución ji-cuadrada

$X \sim \chi^2(n)$ con $n > 0$.

$$f(x) = \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} \quad \text{para } x > 0.$$

$$E(X) = n.$$

$$\text{Var}(X) = 2n.$$

$$M(t) = (1 - 2t)^{-n/2} \quad \text{para } t < 1/2.$$

$$\phi(t) = (1 - 2it)^{-n/2}.$$

Si X tiene distribución $N(0, 1)$, entonces X^2 tiene distribución $\chi^2(1)$.

Distribución log normal

$X \sim \text{log normal}(\mu, \sigma^2)$ con $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp[-(\ln x - \mu)^2 / 2\sigma^2] \quad \text{para } x > 0.$$

$$E(X) = \exp(\mu + \sigma^2/2).$$

$$E(X^n) = \exp(n\mu + n^2\sigma^2/2).$$

$$\text{Var}(X) = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2).$$

La función generadora de momentos no existe. Si X tiene distribución $N(\mu, \sigma^2)$, entonces e^X tiene distribución log normal (μ, σ^2) .

Distribución normal

$$X \sim N(\mu, \sigma^2) \text{ con } \mu \in \mathbb{R} \text{ y } \sigma^2 > 0.$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}.$$

$$E(X) = \mu.$$

$$\text{Var}(X) = \sigma^2.$$

$$M(t) = \exp(\mu t + \sigma^2 t^2/2).$$

$$\phi(t) = \exp(i\mu t - \sigma^2 t^2/2).$$

Cuando $\mu = 0$ y $\sigma^2 = 1$ se obtiene la distribución normal estándar. La suma o diferencia de dos variables independientes con distribución normal tiene distribución normal.

Distribución Pareto

$$X \sim \text{Pareto}(a, b) \text{ con } a > 0 \text{ y } b > 0.$$

$$f(x) = ab^a/(b+x)^{a+1} \text{ para } x > 0.$$

$$F(x) = 1 - [b/(b+x)]^a \text{ para } x > 0.$$

$$E(X) = b/(a-1) \text{ para } a > 1.$$

$$\text{Var}(X) = ab^2/[(a-1)^2(a-2)] \text{ para } a > 2.$$

Distribución Poisson

$X \sim \text{Poisson}(\lambda)$ con $\lambda > 0$.

$f(x) = e^{-\lambda} \lambda^x / x!$ para $x = 0, 1, \dots$

$E(X) = \lambda$.

$\text{Var}(X) = \lambda$.

$G(t) = e^{-\lambda(1-t)}$.

$M(t) = \exp[\lambda(e^t - 1)]$.

La suma de dos variables independientes con distribución $\text{Poisson}(\lambda_1)$ y $\text{Poisson}(\lambda_2)$ tiene distribución $\text{Poisson}(\lambda_1 + \lambda_2)$.

Distribución t

$X \sim t(n)$ con $n > 0$.

$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} (1 + x^2/n)^{-(n+1)/2}$.

$E(X) = 0$.

$\text{Var}(X) = n/(n-2)$ para $n > 2$.

$M(t)$ no existe para $t \neq 0$.

$\phi(t) = \exp(-|t|)$, cuando $n = 1$. La expresión de $\phi(t)$ resulta complicada para valores $n \geq 2$.

Distribución uniforme discreta

$X \sim \text{unif}\{x_1, \dots, x_n\}$ con $n \in \mathbb{N}$.

$f(x) = 1/n$ para $x = x_1, \dots, x_n$.

$E(X) = (x_1 + \dots + x_n)/n$.

$\text{Var}(X) = [(x_1 - \mu)^2 + \dots + (x_n - \mu)^2]/n$.

$G(t) = (t^{x_1} + \dots + t^{x_n})/n$.

$M(t) = (e^{x_1 t} + \dots + e^{x_n t})/n$.

Distribución uniforme continua

$X \sim \text{unif}(a, b)$ con $a < b$.

$f(x) = 1/(b - a)$ para $x \in (a, b)$.

$F(x) = (x - a)/(b - a)$ para $x \in (a, b)$.

$E(X) = (a + b)/2$.

$\text{Var}(X) = (b - a)^2/12$.

$M(t) = (e^{bt} - e^{at})/(bt - at)$.

Distribución Weibull

$X \sim \text{Weibull}(r, \lambda)$ con $r > 0$ y $\lambda > 0$.

$f(x) = e^{-(\lambda x)^r} r \lambda^r x^{r-1}$ para $x > 0$.

$F(x) = 1 - e^{-(\lambda x)^r}$ para $x > 0$.

$E(X) = \Gamma(1 + 1/r)/\lambda$.

$\text{Var}(X) = [\Gamma(1 + 2/r) - \Gamma^2(1 + 1/r)]/\lambda^2$.

Cuando $r = 1$ se obtiene la distribución $\exp(\lambda)$. Cuando $r = 2$ se obtiene la distribución Rayleigh(λ).

APÉNDICE B

Conceptos y resultados varios

El alfabeto griego

A α	alfa	I ι	iota	P ρ, ϱ	rho
B β	beta	K κ	kapa	$\Sigma \sigma, \varsigma$	sigma
$\Gamma \gamma$	gama	$\Lambda \lambda$	lambda	T τ	tau
$\Delta \delta$	delta	M μ	mu	$\Upsilon \upsilon$	upsilon
E ϵ, ε	epsilon	N ν	nu	$\Phi \phi, \varphi$	phi
Z ζ	zeta	$\Xi \xi$	xi	X χ	ji ó chi
H η	eta	O \omicron	omikron	$\Psi \psi$	psi
$\Theta \theta, \vartheta$	theta	$\Pi \pi$	pi	$\Omega \omega$	omega

Notación

- $\mathcal{B}(\mathbb{R})$: Conjuntos de Borel de \mathbb{R} .
 $a \vee b$: $\text{máx}\{a, b\}$.
 $a \wedge b$: $\text{mín}\{a, b\}$.
 $A \perp B$: Independencia de los eventos A y B .
 $[x]$: Parte entera de x .
 $F(x+)$: Límite por la derecha de la función F en el punto x .
 $F(x-)$: Límite por la izquierda de la función F en el punto x .

Lema de Abel

Sea a_0, a_1, \dots una sucesión de números reales o complejos tal que la serie $\sum_{n=0}^{\infty} a_n$ es convergente. Defina la función $G(t) = \sum_{n=0}^{\infty} a_n t^n$, la cual es convergente para valores de t por lo menos en el intervalo $[0, 1]$. El lema de Abel asegura que $G(t)$ es una función continua por la izquierda en $t = 1$, es decir,

$$\lim_{t \nearrow 1} G(t) = \sum_{n=0}^{\infty} a_n.$$

Límite superior e inferior

Sea a_1, a_2, \dots una sucesión infinita de números reales. Para cada m natural defina $b_m = \text{ínf}\{a_m, a_{m+1}, \dots\}$, y $c_m = \text{sup}\{a_m, a_{m+1}, \dots\}$. Claramente $b_m \leq b_{m+1}$, y $c_m \geq c_{m+1}$. Es decir, ambas sucesiones son monótonas, la primera no decreciente y la segunda no creciente, por lo tanto son convergentes, no excluyendo con ello valores infinitos. Al límite de la sucesión $b_1 \leq b_2 \leq \dots$ se le llama *límite inferior*, y al límite de $c_1 \geq c_2 \geq \dots$ se le

llama *límite superior* de la sucesión a_1, a_2, \dots . A estos límites se les denota por $\liminf_{n \rightarrow \infty} a_n$ y $\limsup_{n \rightarrow \infty} a_n$, respectivamente. Es inmediato comprobar que $\liminf_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} a_n$. Además la sucesión original es convergente al número a si, y sólo si, $\liminf_{n \rightarrow \infty} a_n = \limsup_{n \rightarrow \infty} a_n = a$. Estos conceptos de límite inferior y superior pueden extenderse al caso de sucesiones de eventos como se muestra en el primer capítulo de este texto.

Imagen inversa

Sean \mathbb{A} y \mathbb{B} dos conjuntos. Considere una función $X : \mathbb{A} \rightarrow \mathbb{B}$. La imagen inversa de un conjunto $B \subseteq \mathbb{B}$ es un subconjunto de \mathbb{A} , denotado por $X^{-1}B$, y definido como sigue: $X^{-1}B = \{a \in \mathbb{A} : X(a) \in B\}$.

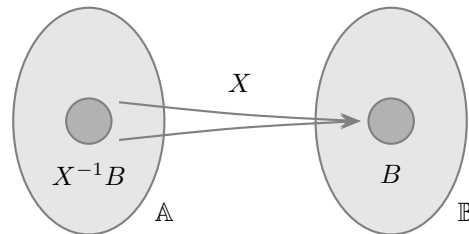


Figura B.1: Imagen inversa.

En palabras, la imagen inversa de B es aquella colección de elementos de \mathbb{A} tal que al aplicarles la función X toman un valor dentro del conjunto B . Observe que X es una función puntual, es decir, lleva puntos de \mathbb{A} en puntos de \mathbb{B} , mientras que X^{-1} es una función conjuntista, es decir, lleva subconjuntos de \mathbb{B} en subconjuntos de \mathbb{A} . No debe confundirse X^{-1} con la función inversa de X .

El concepto de imagen inversa es usado en este texto para definir a una variable aleatoria como una función medible. La imagen inversa cumple las siguientes propiedades:

- a) $X^{-1}\mathbb{B} = \mathbb{A}$.
- b) $X^{-1}(B^c) = (X^{-1}B)^c$.
- c) Si $B_1 \subseteq B_2$, entonces $X^{-1}B_1 \subseteq X^{-1}B_2$.
- d) $X^{-1}(B_2 - B_1) = X^{-1}B_2 - X^{-1}B_1$.
- e) $X^{-1}(\bigcup_{k=1}^{\infty} B_k) = \bigcup_{k=1}^{\infty} X^{-1}B_k$.
- f) $X^{-1}(\bigcap_{k=1}^{\infty} B_k) = \bigcap_{k=1}^{\infty} X^{-1}B_k$.
- g) $X(X^{-1}B) \subseteq B$, la igualdad se cumple si, y sólo si, X es sobre.
- h) $A \subseteq X^{-1}(XA)$, la igualdad se cumple si, y sólo si, X es inyectiva.

Si se tienen dos funciones $X : \mathbb{A} \rightarrow \mathbb{B}$ y $Y : \mathbb{B} \rightarrow \mathbb{C}$, entonces para cualquier subconjunto C de \mathbb{C} , se cumple $(X \circ Y)^{-1}C = X^{-1}(Y^{-1}C)$.

Función indicadora

La función indicadora de un conjunto $A \subseteq \Omega$ es la función $1_A : \Omega \rightarrow \{0, 1\}$ dada por

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

De este modo la función 1_A toma el valor uno dentro del conjunto A , y cero fuera de él. Es sencillo verificar que esta función resulta ser una variable aleatoria si, y sólo si, el conjunto A es un evento. La función indicadora cumple, entre otras, las siguientes propiedades:

- a) $1_{A \cup B} = \max\{1_A, 1_B\} = 1_A + 1_B - 1_A \cdot 1_B$.

- b) $1_{A \cap B} = \min\{1_A, 1_B\} = 1_A \cdot 1_B$.
- c) $1_{A^c} = 1 - 1_A$.
- d) $1_{A-B} = 1_A - 1_A \cdot 1_B$.
- e) $1_{A \Delta B} = |1_A - 1_B| = |1_A - 1_B|^2 = 1_A + 1_B - 2 \cdot 1_A \cdot 1_B$.
- f) Si $A \subseteq B$, entonces $1_A \leq 1_B$.

Esperanza condicional

Sea (Ω, \mathcal{F}) un espacio medible. Sean P y Q dos medidas de probabilidad. Se dice que Q es absolutamente continua respecto de P si cada vez que $P(A) = 0$, necesariamente $Q(A) = 0$ para cada A en \mathcal{F} . En tal caso se escribe $Q \ll P$.

TEOREMA DE RADON-NIKODYM. Si $Q \ll P$, entonces existe una variable aleatoria integrable ξ que es única P -casi seguramente, y es tal que para cada evento A ,

$$Q(A) = \int_A \xi dP.$$

Se escribe $\xi = dQ/dP$ y se le llama la derivada de Radon-Nikodym.

Con ayuda de este teorema es fácil demostrar la existencia y unicidad de la esperanza condicional. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, sea X una variable aleatoria integrable, y sea $\mathcal{G} \subseteq \mathcal{F}$ una sub σ -álgebra. Para cada A en \mathcal{G} defina

$$Q(A) = \int_A X dP.$$

Puede comprobarse que $Q \ll P$ cuando P se restringe a la σ -álgebra \mathcal{G} . El teorema de Radon-Nikodym garantiza entonces la existencia y unicidad P -casi segura de una variable aleatoria \mathcal{G} -medible ξ tal que para cada A en

\mathcal{G} ,

$$\int_A X dP = \int_A \xi dP,$$

A la variable ξ le hemos denotado por $E(X | \mathcal{G})$. He aquí una lista de algunas de sus propiedades.

1. $E(X | \mathcal{G})$ es \mathcal{G} -medible y tiene esperanza finita.
2. $\int_G E(X | \mathcal{G}) dP = \int_G X dP$, para cualquier $G \in \mathcal{G}$.
3. $E(E(X | \mathcal{G})) = E(X)$.
4. $E(X | \{\emptyset, \Omega\}) = E(X)$.
5. Si B es un evento tal que $0 < P(B) < 1$, entonces
 $E(1_A | \{\emptyset, B, B^c, \Omega\}) = P(A | B)1_B + P(A | B^c)1_{B^c}$.
6. Si B_1, \dots, B_n es una partición de Ω tal que cada elemento tiene probabilidad estrictamente positiva, entonces
 $E(X | \sigma\{B_1, \dots, B_n\}) = E(X | B_1) \cdot 1_{B_1} + \dots + E(X | B_n) \cdot 1_{B_n}$.
7. $E(\alpha X + Y | \mathcal{G}) = \alpha E(X | \mathcal{G}) + E(Y | \mathcal{G})$.
8. Si $X \geq 0$, entonces $E(X | \mathcal{G}) \geq 0$.
9. Si $X \leq Y$, entonces $E(X | \mathcal{G}) \leq E(Y | \mathcal{G})$.
10. $|E(X | \mathcal{G})| \leq E(|X| | \mathcal{G})$.
11. $E|E(X | \mathcal{G})| \leq E(|X|)$.
12. CASO DISCRETO. Si Y toma los valores y_1, y_2, \dots con probabilidad estrictamente positiva, entonces $E(X | Y) = \sum_{i=1}^{\infty} E(X | Y = y_i) 1_{(Y=y_i)}$.
13. CASO ABS. CONTINUO. Si ω es tal que $Y(\omega) = y$, entonces

$$E(X | Y)(\omega) = \int_{-\infty}^{\infty} x dF_{X|Y}(x|y), \quad \text{cuando } f_Y(y) \neq 0.$$

14. Si $\mathcal{G}_1 \subseteq \mathcal{G}_2$, entonces $E(E(X | \mathcal{G}_1) | \mathcal{G}_2) = E(E(X | \mathcal{G}_2) | \mathcal{G}_1) = E(X | \mathcal{G}_1)$.
15. Si X es independiente de \mathcal{G} , entonces $E(X | \mathcal{G}) = E(X)$.
16. Si X es \mathcal{G} -medible, entonces $E(X | \mathcal{G}) = X$.
En particular, $E(c | \mathcal{G}) = c$.
17. Si \mathcal{G}_1 y \mathcal{G}_2 son independientes, entonces
 $E(X | \sigma(\mathcal{G}_1 \cup \mathcal{G}_2)) = E(X | \mathcal{G}_1) + E(X | \mathcal{G}_2) - E(X)$.
Si además X es independiente de \mathcal{G}_2 , entonces
 $E(X | \sigma(\mathcal{G}_1 \cup \mathcal{G}_2)) = E(X | \mathcal{G}_1)$.
18. Si $X_n \xrightarrow{m} X$, entonces $E(X_n | \mathcal{G}) \xrightarrow{m} E(X | \mathcal{G})$.
19. TEOREMA DE CONVERGENCIA MONÓTONA. Si $X_n \geq 0$ y $X_n \nearrow X$ c.s., entonces $E(X_n | \mathcal{G}) \nearrow E(X | \mathcal{G})$ c.s.
20. Si XY es integrable y X es \mathcal{G} -medible, entonces $E(XY | \mathcal{G}) = X E(Y | \mathcal{G})$.
21. X es independiente de \mathcal{G} si, y sólo si, $E(f(X) | \mathcal{G}) = E(f(X))$ para cualquier función Lebesgue medible f tal que $f(X)$ es integrable.
22. DESIGUALDAD DE JENSEN. Si u es convexa y $u(X)$ es integrable, entonces $u(E(X | \mathcal{G})) \leq E(u(X) | \mathcal{G})$.

Bibliografía

- [1] Barry R. J. (1981) *Probabilidade: Um Curso em Nível Intermediário*. IMPA-CNPq.
- [2] Blake I. F. (1979) *An Introduction to Applied Probability*. Wiley.
- [3] Bobrowski A. (2005) *Functional Analysis for Probability and Stochastic Processes*. Cambridge University Press.
- [4] Clarke L. E. (1975) *Random Variables*. Longman.
- [5] Cohn D. L. (1980) *Measure Theory*. Birkhäuser.
- [6] Feller W. (1978) *Introducción a la Teoría de las Probabilidades y sus Aplicaciones*. Vol. I y II. Limusa.
- [7] Fristedt B., Gray L. F. (1996) *A Modern Approach to Probability Theory*. Birkhäuser.
- [8] García Álvarez M. A. (2005) *Introducción a la Teoría de la Probabilidad. Primer Curso*. Fondo de Cultura Económica.
- [9] García Álvarez M. A. (2005) *Introducción a la Teoría de la Probabilidad. Segundo Curso*. Fondo de Cultura Económica.
- [10] Grimmett G. R., Stirzaker D. R. (1982) *Probability and Random Processes*. Clarendon Press.

- [11] Grimmett G. R., Stirzaker D. R. (1986) *Probability: an Introduction*. Oxford University Press.
- [12] Grimmett G. R., Stirzaker D. R. (2001) *One Thousand Exercises in Probability*. Oxford University Press.
- [13] Gut A. (2005) *Probability: a Graduate Course*. Springer.
- [14] Halmos P. R. (1960) *Measure Theory*. Van Nostrand.
- [15] Harris B. (1966) *Theory of Probability*. Addison–Wesley.
- [16] Hoel P., Port S., Stone C. (1971) *Probability Theory*. Houghton Mifflin Co.
- [17] Jacod D., Protter P. (2000) *Probability Essentials*. Springer–Verlag.
- [18] Karr A. F. (1993) *Probability*. Springer–Verlag.
- [19] Laha R. G., Rohatgi V. K. (1979) *Probability Theory*. John Wiley & Sons.
- [20] Miller I., Miller M. (1999) *John E. Freund's Mathematical Statistics - 6th ed.* Prentice–Hall.
- [21] Mood A. M., Graybill F. A., Boes D. C. (1974) *Introduction to the Theory of Statistics*. McGraw Hill.
- [22] Parzen E. (1960) *Modern Probability Theory and its Applications*. Wiley.
- [23] Resnick S. I. (1999) *A Probability Path*. Birkhäuser.
- [24] Rincón L. (2004) ¿Qué es la esperanza condicional? *Miscelánea Matemática*. No. **37**, Agosto 2004, SMM.
- [25] Rincón L. (2006) Sobre el problema del mono que escribe caracteres al azar. *Miscelánea Matemática*. No. **42**, SMM.

- [26] Romano J. P. , Siegel A. F. (1986) *Counterexamples in Probability and Statistics*. Chapman & Hall.
- [27] Rosenthal J. S. (2000) *A First Look at Rigorous Probability Theory*. World Scientific.
- [28] Ross S. (1976) *A First Course in Probability*. Prentice Hall.
- [29] Villa Morales J. (2005) *Introducción a la Medida e Integración*. Textos Universitarios, Ciencias Básicas. Universidad Autónoma de Aguascalientes.
- [30] Von Plato J. (1994) *Creating Modern Probability*. Cambridge University Press.
- [31] Williams D. (1991) *Probability with Martingales*. Cambridge University Press.
- [32] Williams D. (2001) *Weighing the Odds: A Course in Probability and Statistics*. Cambridge University Press.

Índice

- σ -álgebra, 3
 - de Borel de \mathbb{R} , 11
 - de Borel de \mathbb{R}^n , 14
 - generada, 7
 - producto, 14
- Álgebra, 9
- Acoplamiento, 159
- Aditividad finita, 24
- Borel-Cantelli, 38
- Cópula, 159
- Cociente
 - de Mills., 140
- Coficiente
 - de correlación, 174
 - multinomial, 182
- Completación de espacios, 27
- Conjunto
 - Borel medible, 11
 - Boreliano, 11
 - de Borel, 11
 - medible, 3
- Continuidad de la prob, 28, 30, 31
- Convergencia
 - casi dondequiera, 290
 - casi segura, 289
 - casi siempre, 290
 - débil, 294
 - de eventos, 16
 - en distribución, 294
 - en media, 293
 - en media cuadrática, 294
 - en probabilidad, 291
 - puntual, 288
- Convolución, 242
- Correlación
 - negativa, 177
 - nula, 177
 - positiva, 177
- Cotas de Fréchet, 188
- Covarianza, 171
- Cuantil
 - de una v.a., 93
- Cuartiles, 93
- Desigualdad
 - c_r , 128
 - de Bonferroni, 53
 - de Boole, 24
 - de Cantelli, 362

- de Cauchy-Schwarz, 126
 - condicional, 227
- de Chebyshev, 349
- de Hölder, 128
- de Jensen, 127
- de Kolmogorov, 350
- de Kounias, 54
- de Markov, 347
 - condicional, 227
- de Minkowski, 129
- Desviación estándar, 90
- Distribución
 - absolutamente continua, 76
 - arcoseno, 138
 - Bernoulli, 95, 365
 - beta, 103, 366
 - binomial, 96, 366
 - binomial negativa, 98, 366
 - bivariada, 144
 - Cauchy, 367
 - continua, 75
 - de acoplamiento, 159
 - discreta, 75
 - exponencial, 101, 367
 - exponencial doble, 135
 - F de Snedecor, 270
 - gama, 101, 368
 - geométrica, 96, 368
 - hipergeométrica, 99, 369
 - multivariada, 182
 - ji-cuadrada, 263, 369
 - log gama, 232
 - log normal, 106, 231, 369
 - multimodal, 94
 - multinomial, 181
 - multivariada, 145
 - normal, 104, 370
 - bivariada, 184
 - estándar, 105
 - multivariada, 185
 - Pareto, 370
 - Poisson, 98, 371
 - Rayleigh, 372
 - singular, 76, 77
 - t de Student, 267, 371
 - trinomial, 182
 - uniforme
 - bivariada, 183
 - continua, 101, 372
 - discreta, 94, 371
 - unimodal, 94
 - univariada, 145
 - Weibull, 372
- Ensayo Bernoulli, 95
- Error
 - absoluto medio, 129
 - cuadrático medio, 123
- Espacio
 - L^1 , 126
 - L^2 , 127
 - de probabilidad, 1, 2
 - completo, 27
 - medible, 3
 - muestral, 2
- Esperanza
 - condicional, 214, 377
 - condicional (evaluada), 162
 - de un vector, 179

- de una función de un vector, 168
- de una función de una v.a., 88
- de una v.a., 85
- Estadística, 261
- Estadísticas de orden, 272
- Evento, 2
 - casi seguro, 80
 - compuesto, 2
 - simple, 2
- Fórmula
 - de inclusión y exclusión, 51
- Función
 - beta, 103
 - Borel medible, 61
 - de acumulación de prob, 68
 - de densidad, 75
 - de masa de probabilidad, 75
 - de probabilidad, 75
 - gama, 102
 - indicadora, 376
 - medible, 112
 - signo, 111
- Función característica, 323
 - fórmula de inversión, 330, 333
 - teorema de continuidad, 334
 - teorema de unicidad, 333
- Función de densidad, 76
 - condicional, 160
 - conjunta, 152
 - marginal, 158
- Función de distribución, 68
 - condicional, 161
 - conjunta, 144
 - marginal, 156
- Función de probabilidad
 - acumulada, 68
 - conjunta, 149
- Función generadora
 - de momentos, 317
 - de momentos factoriales, 316
 - de probabilidad, 311
- Igualdad
 - casi segura, 80
 - en distribución, 80
- Imagen inversa, 375
- Independencia
 - de σ -álgebras, 36
 - de clases, 36
 - de eventos, 34
 - de v.a.s, 163
 - de vectores, 167
- Integral de Riemann-Stieltjes, 81
- Límite inferior
 - de eventos, 15
 - de números, 374
- Límite superior
 - de eventos, 15
 - de números, 374
- Lema de Abel, 374
- Ley de eventos raros, 132
- Ley de los grandes números, 352
 - débil, 352
 - en media cuadrática, 362
 - fuerte, 354
- Matriz
 - de correlación, 181

- de covarianzas, 180
- Media, 85
 - muestral, 262
- Mediana
 - de una v.a., 93
 - muestral, 285
- Medibilidad, 58
- Medida de probabilidad, 2, 20
 - inducida, 58
 - inducida por una v.a., 112
- Moda
 - de una v.a., 94
- Momentos, 92
 - absolutos, 92
 - centrales, 92
 - centrales absolutos, 92
 - factoriales, 92
- Muestra aleatoria, 261
- Paradoja
 - de San Petersburgo, 119
- Probabilidad
 - axiomática, 20
 - clásica, 21
 - condicional, 26
 - frecuentista, 353
 - geométrica, 22
- Problema de los momentos, 92
- Rango
 - de una m.a., 276
- Regla del producto, 53
- Semiálgebra, 10
- Teorema
 - central del límite, 358
 - de Bayes, 27
 - de Bernoulli, 352
 - de cambio de variable, 230, 233, 236
 - de convergencia dominada, 305
 - de convergencia monótona, 303
 - de de Moivre-Laplace, 357
 - de Poisson, 132
 - de probabilidad total, 27
- Valor
 - esperado, 85
 - medio, 85
 - promedio, 85
- Variable aleatoria, 58
 - continua, 75, 76
 - discreta, 75
 - mixta, 77
 - singular, 76, 77
- Varianza
 - condicional, 223
 - de un vector, 179
 - de una v.a., 90
 - muestral, 262
- Vector aleatorio, 141
 - continuo, 143
 - discreto, 143