

Valentín V. Petrov

Ernesto Mordecki

TEORÍA DE LA PROBABILIDAD

Facultad de Ciencias

DIRAC – 2008/2023

Los conceptos vertidos en los libros editados por la Facultad de Ciencias de la Universidad de la República, son de responsabilidad de sus autores. Su edición no implica que dichos conceptos sean compartidos por las mencionadas instituciones.

*La publicación de este libro fue realizada con el apoyo de la
Comisión Sectorial de Investigación Científica (CSIC)
de la Universidad de la República.*

Petrov, Valentín

*Teoría de la Probabilidad / Valentín Vladímirovich Petrov,
Ernesto Mordecki Pupko. – 2a. ed. –
Montevideo: DIRAC, 2008/2023.*

272 pp. : 12 il.

Bibliografía p. 267

ISBN: 978-9974-0-0433-7

1. PROBABILIDAD 2. PROCESOS ESTOCÁSTICOS

60-01

AMS MSC2000

Imagen de tapa: “Distribuciones Gaussianas II” de Anatoly T. Fomenko
(cortesía del autor)

Diseño de tapa: Alejandro Crosa

Asistente de edición: Gabriel Santoro

Publicado por DIRAC – Facultad de Ciencias – Universidad de la República

Calle Iguá 4225 casi Mataojo – Montevideo – Uruguay

Tel. (0598 2) 525 17 11 – Fax (0598 2) 525 86 17 – e-mail: dirac@fcien.edu.uy

© de la primera edición: Editorial URSS, 2002.

© de la segunda edición: DIRAC – Facultad de Ciencias, 2008.

A Valentina y Rosana

Prólogo

La literatura dedicada a la enseñanza de la teoría de la probabilidad es muy extensa; existen numerosos libros de texto, excelentemente escritos, para lectores con diferentes niveles de formación en matemática. Entre estos textos, podemos destacar los escritos por Borovkov [1], Feller [2], Gnedenko [3], Gut [4], Ross [6], y Shiryaev [7], incluidos en la bibliografía al final de este libro. Sin embargo, la literatura dedicada a esta temática en idioma español es escasa, y tenemos la esperanza de que la presente publicación llenará este vacío, en alguna medida.

Este libro contiene un primer curso de teoría de la probabilidad, basado en cursos dictados por ambos autores, a lo largo de muchos años, en la Universidad de San Petersburgo (Rusia) y en la Universidad de la República (Montevideo, Uruguay).

En el proceso de su preparación se han tenido en cuenta especialmente los intereses de lectores con diferentes niveles de preparación matemática: el material contenido en el libro es perfectamente accesible para quienes hayan estudiado los temas de un curso habitual de cálculo diferencial e integral. Los lectores en esta situación, podrán restringirse a la consideración de variables aleatorias con distribuciones discretas o absolutamente continuas, que son las encontradas en las aplicaciones; se presta especial atención a estas dos clases de distribuciones. En particular, se presenta una exposición detallada de las nociones de esperanza matemática de una variable aleatoria, varianza de una variable aleatoria, esperanza condicional de una variable aleatoria con respecto de otra, y cuestiones relacionadas, para estas dos clases de distribuciones.

Al mismo tiempo, y en forma independiente, se definen estas nociones en los términos habituales de la teoría de la medida e integración con respecto de medidas abstractas¹. Esta segunda exposición está dirigida a

¹El lector interesado en tomar contacto con los elementos básicos de la teoría de la

estudiantes de matemática o estadística, quienes encontrarán una presentación rigurosa, de interés, y actualizada de la disciplina.

Cada capítulo se acompaña de un conjunto de ejercicios, ordenados según su grado de dificultad, y el lector no debe desanimarse si no resuelve todos los ejercicios. Se prestó especial cuidado en la presentación de las demostraciones de los teoremas, proposiciones y lemas, por lo que este libro puede utilizarse en forma autodidacta.

La iniciativa de realizar el presente libro correspondió a V. Petrov, quien escribió los 7 primeros capítulos. E. Mordecki escribió los últimos tres capítulos, y preparó el texto en español. Todo el material fue discutido y revisado en forma conjunta.

Varias personas estuvieron involucradas, en diferentes formas, con la preparación de este libro. Walter Moreira preparó los gráficos y las tablas, y prestó invaluable ayuda en la preparación de la versión electrónica; Ricardo Fraiman e Isabel Cañete leyeron partes del manuscrito, sugiriendo mejoras y correcciones. A ellos nuestro agradecimiento. Un especial reconocimiento merecen Valentina y Rosana por su aliento, paciencia, y comprensión.

Este libro fue posible gracias al apoyo del Centro de Matemática, la Comisión Sectorial de Investigación Científica (CSIC), y el Laboratorio de Probabilidad y Estadística, en la Universidad de la República; junto con el PEDECIBA–Matemática, y es el resultado de la colaboración científica entre nuestros países; tenemos la esperanza de que ayude a su fortalecimiento.

Los autores esperan que su trabajo resulte de utilidad a aquellas personas que estudian o enseñan teoría de la probabilidad.

Montevideo, abril de 2002.

V. Petrov, E. Mordecki.

En la presente segunda edición se han corregido algunas erratas y agregado las soluciones de algunos ejercicios. Esta edición es posible gracias al apoyo del programa de publicaciones de la Comisión Sectorial de Investigación Científica de la Universidad de la República (CSIC).

Montevideo, mayo de 2008.

V. Petrov, E. Mordecki.

medida, y de la integración con respecto de medidas abstractas (análisis real), luego de un curso de cálculo diferencial e integral, podrá utilizar la excelente exposición en el libro de Borovkov [1].

Introducción para quienes comienzan a estudiar teoría de la probabilidad

La teoría de la probabilidad estudia modelos matemáticos de fenómenos aleatorios. Los fenómenos *aleatorios* son aquellos en los que la verificación de un cierto conjunto de condiciones determinadas, conduce a un resultado de una serie de resultados posibles. Por contraposición, los fenómenos *determinísticos*, o no aleatorios, son aquellos en los que la verificación de un cierto conjunto de condiciones determinadas conduce, en forma inevitable, a un resultado fijo (por ejemplo: el enfriamiento del agua hasta 0 grados centígrados bajo presión atmosférica normal conduce a la formación del hielo).

Llamamos *experimento* a cada uno de esos conjuntos de condiciones determinadas. Si un experimento consiste, por ejemplo, en tirar una moneda al aire, cada realización de este experimento conduce a uno de dos resultados posibles: la aparición de cara o la aparición de número, y no podemos, antes de realizar el experimento, conocer el resultado. Sin embargo, al repetir este experimento en una serie una gran cantidad de veces, resulta que la cantidad de veces que aparece cara es, aproximadamente, la mitad de la cantidad de experimentos que componen la serie.

Dado un cierto experimento, llamamos *suceso* a cada uno de sus resultados posibles, y utilizamos las letras $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ (con índices o sin ellos) para designar los sucesos.

La *frecuencia* de un suceso \mathbf{A} en un serie de n experimentos (o *frecuencia relativa*), se define como la proporción $f(\mathbf{A}) = m/n$, donde m es la cantidad de experimentos en los que el ocurrió el suceso \mathbf{A} . Es fácil de ver que la frecuencia así definida verifica las siguientes propiedades: (1) $0 \leq f(\mathbf{A}) \leq 1$ para cualquier suceso \mathbf{A} ; (2) $f(\Omega) = 1$, si Ω representa

el suceso *cierto* (es decir, un suceso que ocurre indefectiblemente en cada experimento); (3) $f(\mathbf{A} \text{ ó } \mathbf{B}) = f(\mathbf{A}) + f(\mathbf{B})$, si los sucesos \mathbf{A} y \mathbf{B} son incompatibles (es decir, no pueden ocurrir ambos sucesos simultáneamente).

Existe un muy amplio conjunto de situaciones en las que tiene lugar la estabilidad de las frecuencias antes mencionada; mas precisamente, en la verificación de un cierto conjunto de condiciones determinadas n_1 veces, luego n_2 veces, \dots , luego n_k veces (es decir, se están llevando a cabo series de experimentos compuestas cada una por n_1, n_2, \dots, n_k experimentos), las frecuencias de un suceso fijo \mathbf{A} , resultante en las diferentes series, serán muy cercanas entre sí, siendo además esta proximidad mayor, en general, cuanto mayores sean los largos n_1, n_2, \dots, n_k de estas series de experimentos.

Si el experimento consiste, por ejemplo, en tirar un dado (aquí tenemos 6 resultados posibles, correspondientes a la aparición en la cara superior del dado de una cantidad de puntos igual a 1,2,3,4,5 ó 6), se observa, luego de llevar a cabo algunas series de experimentos prolongadas, que la frecuencia resultante, por ejemplo, de la aparición de 6 puntos en cada serie, será aproximadamente igual a $1/6$. Existe también estabilidad en las frecuencias en indicadores de calidad de un cierto artículo, que se produce en serie. El porcentaje de artículos fallados, encontrado para distintas muestras de gran tamaño en la producción del artículo considerado, habitualmente, resulta prácticamente constante.

Esta constante, en torno a la cual se presentan las fluctuaciones de la frecuencia de un suceso \mathbf{A} considerado, cuando se llevan a cabo series de experimentos prolongadas, se denomina *probabilidad* del suceso \mathbf{A} . De esta forma, la probabilidad de un suceso \mathbf{A} se puede considerar el valor teórico (o ideal) de la frecuencia de este suceso. La relación entre el concepto teórico de probabilidad y el concepto empírico de frecuencia, es como la relación entre una magnitud física (por ejemplo, la longitud de una mesa) y los resultados de su medición.

Lo dicho hasta ahora no es suficiente para construir una teoría matemática de los fenómenos aleatorios. La teoría de la probabilidad es parte de la matemática, y al igual que otras teorías, como por ejemplo la geometría, se construye sobre la base de un sistema de axiomas. La elección del sistema de axiomas se puede realizar de distintas formas. En el comienzo del capítulo 1 se expone un sistema de axiomas que en conjunto define la noción de probabilidad de forma tal, que las reglas válidas para las probabilidades coinciden con las reglas de las frecuencias antes descritas.

En la construcción de este sistema de axiomas dejamos de lado toda otra propiedad de las frecuencias, y asumimos la idea intuitiva de que la probabilidad de un suceso es el valor teórico de la frecuencia de este suceso.

Índice general

Prólogo	5
Para quienes comienzan a estudiar Probabilidad	7
1. Conceptos básicos	15
1.1. Sucesos	15
1.2. Axiomas de la teoría de la probabilidad	16
1.3. Primeras consecuencias de los axiomas	20
1.4. Regla clásica del cálculo de probabilidades	23
1.5. Probabilidad condicional. Probabilidad total y Bayes.	25
1.6. Sucesos independientes	29
1.7. Ejercicios	31
2. Esquema de Bernoulli	37
2.1. Esquema de Bernoulli y fórmula de la distribución binomial	37
2.2. Teorema límite local de De Moivre–Laplace	41
2.3. Teorema límite integral de De Moivre–Laplace	44
2.4. Teorema de Bernoulli	50
2.5. Aproximación de Poisson a la distribución binomial	52
2.6. Ejercicios	53
3. Variables aleatorias y distribuciones de probabilidad	57
3.1. Variables aleatorias y funciones de distribución	57
3.2. Variables aleatorias discretas y absolutamente continuas	61
3.3. Vectores aleatorios y variables aleatorias independientes.	71
3.4. Distribución de la suma de variables aleatorias independientes	80
3.5. Ejercicios	82

4. Esperanza, varianza, y otros momentos	87
4.1. Esperanza matemática	87
4.2. Varianza	97
4.3. Desigualdad de Chebishev	102
4.4. Momentos de órdenes superiores. Mediana y cuantiles . . .	103
4.5. Covarianza, coeficiente de correlación. Matriz de Covarianza	106
4.6. Ejercicios	111
5. Distintas convergencias. Ley de los grandes números	117
5.1. Distintos tipos de convergencia en teoría de la probabilidad.	117
5.2. Ley de los grandes números	121
5.3. Ejercicios	132
6. Funciones características	135
6.1. Definiciones y primeras propiedades	135
6.2. Fórmula de inversión. Teorema de unicidad	142
6.3. Teoremas de Helly	146
6.4. Convergencia de distribuciones y funciones características .	150
6.5. Ejercicios	152
7. Teorema central del límite	157
7.1. Teorema de Lindeberg–Lévy	157
7.2. Teorema de Lindeberg	161
7.3. Teorema de Lyapunov	166
7.4. Ejercicios	168
8. Cadenas de Markov	171
8.1. Definiciones	171
8.2. Clasificación de estados. Estados esenciales y periódicos . .	177
8.3. Recurrencia	181
8.4. Probabilidades límites y distribuciones estacionarias	191
8.5. Ejercicios	201
9. Martingalas	205
9.1. Esperanza condicional	205
9.2. Propiedades de la esperanza condicional	211
9.3. Martingalas	217
9.4. Teorema del muestreo opcional	221
9.5. Convergencia de martingalas	225

	13
9.6. Ley fuerte de los grandes números	230
9.7. Ejercicios	231
10. Proceso de Poisson y proceso de Wiener	237
10.1. Proceso de Poisson. Definición y caracterizaciones	238
10.2. Proceso de Poisson compuesto y aplicaciones	246
10.3. Proceso de Wiener. Definición y primeras propiedades	250
10.4. Problemas de barrera para el proceso de Wiener	253
10.5. Ejercicios	259
Soluciones de algunos ejercicios	263
Tabla de la distribución normal estándar	265
Tabla de la densidad normal estándar	266
Bibliografía	267

Capítulo 1

Conceptos básicos

1.1. Sucesos

Consideremos un cierto conjunto Ω no vacío, que llamamos *espacio de sucesos elementales*. A sus elementos, que llamamos *sucesos elementales* o *puntos*, los designamos con la letra ω , con índices o sin ellos.

Sea \mathcal{A} un cierto conjunto no vacío de subconjuntos de Ω , que cumple las siguientes propiedades: (1) si $\mathbf{A} \in \mathcal{A}$ entonces $\Omega \setminus \mathbf{A} \in \mathcal{A}$; (2) si $\mathbf{A}_1, \mathbf{A}_2, \dots$ es un conjunto finito o numerable de subconjuntos pertenecientes a \mathcal{A} , entonces $\bigcup_n \mathbf{A}_n \in \mathcal{A}$. El conjunto \mathcal{A} se llama σ -álgebra de sucesos, o *campo boreliano de sucesos*, y sus elementos se llaman *sucesos*.

Observemos que el conjunto de todos los subconjuntos de un espacio Ω es una σ -álgebra de sucesos, pero no toda σ -álgebra de sucesos es el conjunto de todos los subconjuntos de algún espacio Ω .

Si \mathcal{A} es una σ -álgebra de sucesos, tenemos $\Omega \in \mathcal{A}$ en vista de la igualdad $\mathbf{A} \cup (\Omega \setminus \mathbf{A}) = \Omega$, válida para todo conjunto \mathbf{A} ; además, el conjunto vacío \emptyset (llamado *suceso imposible*) también pertenece a \mathcal{A} . Si $\mathbf{A}_1, \mathbf{A}_2, \dots$ es un conjunto finito o numerable de sucesos pertenecientes a la σ -álgebra \mathcal{A} , entonces $\bigcap_n \mathbf{A}_n \in \mathcal{A}$ en vista de la igualdad $(\bigcap_n \mathbf{A}_n)^c = \bigcup_n \mathbf{A}_n^c$, donde $\mathbf{B}^c = \Omega \setminus \mathbf{B}$ (complemento del conjunto \mathbf{B}) para cualquier conjunto \mathbf{B} .

En resumen, toda σ -álgebra de sucesos es un conjunto de subconjuntos (no necesariamente todos) de un espacio de sucesos elementales Ω , que contiene, junto con cada uno de sus elementos a su complemento, y junto con cualquier conjunto finito o numerable de sus elementos a su unión y a su intersección; además, el propio espacio de sucesos elementales Ω y el

conjunto vacío \emptyset pertenecen a toda σ -álgebra de sucesos.

El surgimiento de la teoría de la probabilidad es muy anterior a la creación de la teoría de conjuntos. Por ésto, desde su mismo inicio, en teoría de la probabilidad se utilizó (y continúa utilizándose) una terminología específica, diferente de la terminología utilizada en teoría de conjuntos. En la página 17 se presenta una tabla de términos teórico-conjuntistas, junto con los correspondientes términos teórico-probabilistas, que utilizamos a lo largo de este libro. Las letras $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$, con índices o sin ellos, designan a los sucesos, es decir, a los elementos de una σ -álgebra de sucesos \mathcal{A} , relacionada con algún espacio de sucesos elementales Ω .

1.2. Axiomas de la teoría de la probabilidad

Consideremos un espacio de sucesos elementales Ω , y una cierta σ -álgebra de sucesos \mathcal{A} . Mediante las letras $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ (con índices o sin ellos) en lo que sigue designamos a los sucesos, es decir, a los elementos de la σ -álgebra de sucesos \mathcal{A} . Las tres proposiciones siguientes componen el sistema de axiomas de la teoría de la probabilidad:

Axioma I. A cada suceso \mathbf{A} le corresponde un número no negativo $\mathbf{P}(\mathbf{A})$, llamado *probabilidad* del suceso \mathbf{A} .

Axioma II. $\mathbf{P}(\Omega) = 1$.

Axioma III. Si $\mathbf{A}_1, \mathbf{A}_2, \dots$ es un conjunto finito o numerable de sucesos incompatibles dos a dos, entonces

$$\mathbf{P}\left(\bigcup_i \mathbf{A}_i\right) = \sum_i \mathbf{P}(\mathbf{A}_i).$$

Este sistema de axiomas fue propuesto por A. N. Kolmogorov en 1933 y es el utilizado en la actualidad.

En el lenguaje del análisis real, una probabilidad \mathbf{P} es una función de conjunto numerablemente aditiva y no negativa (es decir, una medida positiva), que cumple la condición adicional $\mathbf{P}(\Omega) = 1$.

La terna $(\Omega, \mathcal{A}, \mathbf{P})$, donde \mathbf{P} es una probabilidad definida para cada elemento de la σ -álgebra \mathcal{A} (es decir, para cualquier suceso) que verifica el sistema de axiomas propuesto, se llama *espacio de probabilidad*. En análisis real, un espacio de probabilidad es un espacio medible (Ω, \mathcal{A}) con una medida no negativa \mathbf{P} , que verifica $\mathbf{P}(\Omega) = 1$.

<i>Notación</i>	<i>Término de la teoría de conjuntos</i>	<i>Término de la teoría de la probabilidad</i>
Ω	espacio de elementos	espacio de sucesos elementales
\emptyset	conjunto vacío	suceso imposible
$A \cup B$	unión de los conjuntos A y B	suma de los sucesos A y B
$A \cap B, AB$	intersección de los conjuntos A y B	producto de los sucesos A y B
$AB = \emptyset$	los conjuntos A y B son disjuntos (no tienen elementos comunes)	los sucesos A y B son incompatibles
$C = AB$	el conjunto C es la intersección de los conjuntos A y B	el suceso C consiste en la ocurrencia (simultánea) de ambos sucesos A y B
$D = A \cup B$	el conjunto D es la unión de los conjuntos A y B	el suceso D consiste en la ocurrencia de al menos uno de los sucesos A ó B
$A_i A_j = \emptyset$ ($i, j = 1, 2, \dots;$ $i \neq j$)	los conjuntos A ₁ , A ₂ , ... son disjuntos dos a dos	los sucesos A ₁ , A ₂ , ... son incompatibles dos a dos
$\bigcup_{i=1}^n A_i = \Omega$	cada punto del espacio Ω pertenece por lo menos a uno de los conjuntos A ₁ , ..., A _n	alguno de los sucesos A ₁ , ..., A _n ocurre
$A \subset B$	el conjunto A está contenido en el conjunto B	la ocurrencia del suceso A implica la ocurrencia del suceso B
$\Omega \setminus A$	complemento del conjunto A (designado A ^c)	suceso contrario al suceso A (designado \bar{A})

Ejemplo 1.1. Consideremos un espacio de sucesos elementales Ω compuesto por dos puntos ω_1 y ω_2 . Llamamos *éxito* al punto ω_1 , y fracaso al punto ω_2 . Este espacio de sucesos elementales describe el conjunto de resultados de un experimento, que puede concluir solamente con uno de dos resultados posibles. Un experimento de este tipo es, por ejemplo, el consistente en arrojar por única vez una moneda al aire. Dicho experimento puede concluir solamente con uno de dos resultados posibles: la aparición de cara, o la aparición de número.

Sea \mathcal{A} el conjunto de todos los subconjuntos del espacio de sucesos elementales Ω . (Este conjunto es automáticamente una σ -álgebra.) El conjunto \mathcal{A} está compuesto, en el ejemplo dado, por los cuatro elementos siguientes: \emptyset , $\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_1, \omega_2\} = \Omega$. Asignemos a estos cuatro sucesos las probabilidades siguientes: $0, 1/2, 1/2$ y 1 . Es sencillo ver que las probabilidades así definidas verifican todos los axiomas.

Es importante observar, que la forma indicada de introducir las probabilidades de los sucesos no es única. Más precisamente, si asignamos a los cuatro sucesos considerados los números $0, p, q$ y 1 como probabilidades, donde p y q son números no negativos que verifican la condición $p + q = 1$, también se verifican todos los axiomas. La primer forma de asignar las probabilidades a los sucesos es un caso particular de la segunda, en la que $p = q = 1/2$, y puede utilizarse en la construcción de un modelo matemático para un experimento consistente en arrojar, por única vez, una moneda equilibrada al aire; en esta situación ninguna de las dos caras de la moneda tiene ventaja objetiva sobre la otra, y podemos considerar idénticas las probabilidades de aparición de cara y número. Si la moneda está desequilibrada, o adulterada, la propiedad mencionada de igualdad para la aparición de una de las caras de la moneda (es decir, la equiprobabilidad de ambos resultados) puede no cumplirse, y resulta mas adecuado el segundo método de introducir probabilidades, en el que se le asignan probabilidades de aparición distintas a la cara y al número.

Sin duda, la no unicidad en la forma de introducir las probabilidades de los sucesos resulta una ventaja del sistema de axiomas, dado que permite una mayor flexibilidad en la construcción de modelos matemáticos de fenómenos aleatorios, que tengan en cuenta las especificidades de estos fenómenos.

Ejemplo 1.2. Consideremos un espacio de sucesos elementales Ω compuesto por n puntos $\omega_1, \omega_2, \dots, \omega_n$. Este espacio de sucesos elementales describe el conjunto de resultados posibles de un experimento, que puede

concluir solamente con uno de n resultados. Un experimento de este tipo es, por ejemplo, el consistente en tirar, por única vez, un dado. Este experimento concluye al caer el dado, con la aparición en su cara superior de un número de puntos igual a 1, 2, 3, 4, 5 ó 6. De esta forma, para este experimento, tenemos $n = 6$.

Sea \mathcal{A} el conjunto de todos los subconjuntos del espacio de sucesos elementales Ω . El conjunto \mathcal{A} está formado por los siguientes elementos: el conjunto vacío \emptyset ; n conjuntos de un único elemento $\{\omega_1\}, \dots, \{\omega_n\}$; C_2^n conjuntos de dos elementos $\{\omega_1, \omega_2\}, \dots, \{\omega_1, \omega_n\}, \{\omega_2, \omega_3\}, \dots, \{\omega_2, \omega_n\}, \dots, \{\omega_{n-1}, \omega_n\}$; C_3^n conjuntos de tres elementos $\{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \dots, \{\omega_{n-2}, \omega_{n-1}, \omega_n\}$; \dots ; y $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, un único conjunto de n elementos. Observemos que la cantidad de elementos del conjunto \mathcal{A} es igual a $1 + n + C_2^n + \dots + C_n^n = (1 + 1)^n = 2^n$.

Supongamos que un suceso \mathbf{A} está compuesto por k puntos del espacio de sucesos elementales Ω , es decir

$$\mathbf{A} = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\},$$

donde i_1, i_2, \dots, i_k son k números distintos dos a dos, elegidos entre los naturales $1, 2, \dots, n$. Asignamos al suceso \mathbf{A} una probabilidad igual a la suma de las probabilidades de los sucesos elementales que lo componen, es decir, ponemos

$$\mathbf{P}(\mathbf{A}) = \sum_{m=1}^k \mathbf{P}(\omega_{i_m}),$$

mientras que definimos las probabilidades de cada uno de los sucesos elementales del espacio la siguiente forma: $\mathbf{P}(\omega_i) = p_i$ ($i = 1, \dots, n$), donde p_1, \dots, p_n son números no negativos, y tales que $p_1 + \dots + p_n = 1$. Es fácil de ver que las probabilidades así definidas verifican todos los axiomas. De este modo, hemos construido un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, que puede considerarse como el modelo matemático de un experimento con n resultados posibles.

Si dicho experimento consiste en tirar, por única vez, un dado equilibrado ($n = 6$), podemos asignar como probabilidad del resultado, correspondiente a la aparición de i puntos ($i = 1, \dots, 6$), el número $1/6$. Si se trata de un dado intencionalmente adulterado, falso, resulta más adecuado otra elección de probabilidades, que considere la disparidad entre las distintas caras del dado.

Ejemplo 1.3. Consideremos el espacio de sucesos elementales compuesto por los puntos del intervalo cerrado $[0, 1]$. Sea \mathcal{A} la σ -álgebra de los conjuntos borelianos de este intervalo¹. Sea además \mathbf{P} la medida de Lebesgue. La terna $(\Omega, \mathcal{A}, \mathbf{P})$ es un espacio de probabilidad. Observemos que para cada intervalo $\mathbf{I} = (a, b)$ contenido en $[0, 1]$, tenemos $\mathbf{P}(\mathbf{I}) = b - a$, es decir, la probabilidad del intervalo (a, b) coincide con la longitud de este intervalo.

1.3. Primeras consecuencias de los axiomas

Sea $(\Omega, \mathcal{A}, \mathbf{P})$ un espacio de probabilidad. Las letras $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$ (con índices o sin ellos) designan sucesos, es decir, elementos de la σ -álgebra de sucesos \mathcal{A} .

Propiedad 1. *Para cualquier suceso \mathbf{A} se tiene $\mathbf{P}(\bar{\mathbf{A}}) = 1 - \mathbf{P}(\mathbf{A})$.*

Demostración. Por definición de $\bar{\mathbf{A}}$ (suceso contrario al suceso \mathbf{A}), tenemos $\bar{\mathbf{A}} = \Omega \setminus \mathbf{A}$. De aquí resulta $\mathbf{A}\bar{\mathbf{A}} = \emptyset$. Como por el axioma II tenemos $\mathbf{P}(\mathbf{A} \cup \bar{\mathbf{A}}) = \mathbf{P}(\Omega) = 1$, aplicando el axioma III concluimos, que $1 = \mathbf{P}(\Omega) = \mathbf{P}(\mathbf{A} \cup \bar{\mathbf{A}}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\bar{\mathbf{A}})$. \square

Propiedad 2. *El suceso imposible tiene probabilidad nula: $\mathbf{P}(\emptyset) = 0$.*

Demostración. Esta igualdad se obtiene de la propiedad anterior, si consideramos $\mathbf{A} = \Omega$ y aplicamos el axioma II. \square

Propiedad 3. *Si $\mathbf{A} \subset \mathbf{B}$, entonces $\mathbf{P}(\mathbf{A}) \leq \mathbf{P}(\mathbf{B})$.*

Demostración. Como $\mathbf{A} \subset \mathbf{B}$, tenemos $\mathbf{B} = \mathbf{A} \cup (\mathbf{B} \setminus \mathbf{A})$. Es claro que $\mathbf{B} \setminus \mathbf{A} = \mathbf{B}\bar{\mathbf{A}}$ es un suceso, además los sucesos \mathbf{A} y $\mathbf{B} \setminus \mathbf{A}$ son incompatibles. Por los axiomas III y I tenemos $\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B} \setminus \mathbf{A}) \geq \mathbf{P}(\mathbf{A})$. \square

Propiedad 4. *Para cualquier suceso \mathbf{A} se tiene $0 \leq \mathbf{P}(\mathbf{A}) \leq 1$.*

Demostración. En virtud del axioma I es suficiente demostrar la segunda desigualdad, la cual se deduce inmediatamente de la propiedad anterior, por la inclusión $\mathbf{A} \subset \Omega$ y el axioma II. \square

¹La clase de los subconjuntos borelianos de un cierto intervalo \mathbf{J} es la mínima σ -álgebra de conjuntos de puntos del intervalo \mathbf{J} , que contiene a todos los subintervalos de \mathbf{J} .

Propiedad 5. Para sucesos \mathbf{A} y \mathbf{B} arbitrarios vale la igualdad

$$\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B}). \quad (1.1)$$

Demostración. Es claro que $\mathbf{A} = \mathbf{AB} \cup \mathbf{A}\bar{\mathbf{B}}$ y $\mathbf{B} = \mathbf{AB} \cup \bar{\mathbf{A}}\mathbf{B}$. Como los sucesos en las dos sumas anteriores son incompatibles, por el axioma III, resulta

$$\mathbf{P}(\mathbf{A}) = \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{A}\bar{\mathbf{B}}), \quad \mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\bar{\mathbf{A}}\mathbf{B}).$$

Tenemos también $\mathbf{A} \cup \mathbf{B} = \mathbf{AB} \cup \mathbf{A}\bar{\mathbf{B}} \cup \bar{\mathbf{A}}\mathbf{B}$, donde a la derecha se suman tres sucesos incompatibles dos a dos. Aplicando nuevamente el axioma III y sustituyendo, resulta

$$\begin{aligned} \mathbf{P}(\mathbf{A} \cup \mathbf{B}) &= \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{A}) - \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{AB}) \\ &= \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) - \mathbf{P}(\mathbf{A} \cap \mathbf{B}), \end{aligned}$$

que es la igualdad buscada. \square

Observación. Si los sucesos \mathbf{A} y \mathbf{B} son incompatibles, entonces $\mathbf{P}(\mathbf{AB}) = 0$, y de la fórmula (1.1) se obtiene la igualdad ya conocida $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B})$.

Observación. En forma análoga, no es difícil demostrar, que para tres sucesos \mathbf{A} , \mathbf{B} y \mathbf{C} arbitrarios, tiene lugar la igualdad

$$\begin{aligned} \mathbf{P}(\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}) &= \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B}) + \mathbf{P}(\mathbf{C}) \\ &\quad - \mathbf{P}(\mathbf{AB}) - \mathbf{P}(\mathbf{AC}) - \mathbf{P}(\mathbf{BC}) + \mathbf{P}(\mathbf{ABC}). \end{aligned}$$

Es posible también demostrar una fórmula general: para sucesos arbitrarios $\mathbf{A}_1, \dots, \mathbf{A}_n$ vale la igualdad

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) &= \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(\mathbf{A}_i \mathbf{A}_j) + \sum_{1 \leq i < j < k \leq n} \mathbf{P}(\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k) \\ &\quad - \dots + (-1)^{n+1} \mathbf{P}(\mathbf{A}_1 \dots \mathbf{A}_n). \end{aligned}$$

De aquí es posible obtener la *desigualdad de Bonferroni*:

$$\mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) \geq \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) - \sum_{1 \leq i < j \leq n} \mathbf{P}(\mathbf{A}_i \mathbf{A}_j).$$

Propiedad 6. *Dados n sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ arbitrarios, tiene lugar la desigualdad*

$$\mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) \leq \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i).$$

Demostración. Para $n = 2$ esta desigualdad se obtiene de (1.1). Para $n > 2$, es fácil obtener el resultado anterior, mediante la aplicación sucesiva de la desigualdad $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) \leq \mathbf{P}(\mathbf{A}) + \mathbf{P}(\mathbf{B})$. \square

Propiedad 7. *Sean $\mathbf{A}_1, \dots, \mathbf{A}_n$ sucesos incompatibles dos a dos, y tales que alguno de ellos ocurre². Entonces $\sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) = 1$.*

Demostración. Tenemos $\mathbf{A}_i \mathbf{A}_j = \emptyset$ cuando $i \neq j$, para $i, j = 1, \dots, n$. Además $\bigcup_{i=1}^n \mathbf{A}_i = \Omega$. De los axiomas II y III obtenemos

$$1 = \mathbf{P}(\Omega) = \mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\right) = \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i),$$

concluyendo la demostración. \square

Propiedad 8. *Sea $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \mathbf{A}_3 \supset \dots$ una sucesión de sucesos, y designemos $\mathbf{A} = \bigcap_{i=1}^{\infty} \mathbf{A}_i$. Entonces, existe el $\lim_n \mathbf{P}(\mathbf{A}_n)$, y es igual a $\mathbf{P}(\mathbf{A})$.*

Demostración. Para cada n , tenemos

$$\mathbf{A}_n = \bigcup_{k=n}^{\infty} (\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \cup \mathbf{A}.$$

Como los sucesos que aparecen a la derecha son incompatibles dos a dos, utilizando el axioma III, obtenemos

$$\mathbf{P}(\mathbf{A}_n) = \sum_{k=n}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) + \mathbf{P}(\mathbf{A}). \quad (1.2)$$

Si aquí tomamos $n = 1$, resulta

$$\sum_{k=1}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \leq \mathbf{P}(\mathbf{A}_1) \leq 1.$$

²Ver tabla en la página 17.

De la convergencia de la serie a la izquierda en la fórmula anterior, surge que

$$\sum_{k=n}^{\infty} \mathbf{P}(\mathbf{A}_k \setminus \mathbf{A}_{k+1}) \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

Tomando límite en ambos lados de la igualdad (1.2), obtenemos

$$\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = \mathbf{P}(\mathbf{A}),$$

concluyendo la demostración. \square

1.4. Regla clásica del cálculo de probabilidades

Consideremos un espacio de sucesos elementales Ω compuesto por n puntos $\omega_1, \dots, \omega_n$. Sea \mathcal{A} el conjunto de todos los subconjuntos del espacio Ω . Si $\mathbf{A} = \{\omega_{i_1}, \dots, \omega_{i_k}\}$, asignamos

$$\mathbf{P}(\mathbf{A}) = \sum_{m=1}^k \mathbf{P}(\omega_{i_m}), \quad (1.3)$$

en donde asignamos además

$$\mathbf{P}(\omega_1) = \mathbf{P}(\omega_2) = \dots = \mathbf{P}(\omega_n) = 1/n. \quad (1.4)$$

De esta forma, introducimos las probabilidades de los sucesos como en el ejemplo 1.2, pero eligiendo las probabilidades de los puntos $\omega_1, \dots, \omega_n$ iguales entre sí.

Los puntos $\omega_{i_1}, \dots, \omega_{i_k}$ se llaman *casos favorables* para la ocurrencia del suceso \mathbf{A} , mientras que cada punto ω_i ($i = 1, \dots, n$) es un *caso posible*. De las fórmulas (1.3) y (1.4), resulta que

$$\mathbf{P}(\mathbf{A}) = k/n, \quad (1.5)$$

es decir, la probabilidad del suceso \mathbf{A} es la razón entre el número de casos favorables y el número total de casos posibles.

Es importante observar que la igualdad (1.5) se obtiene como una consecuencia del sistema de axiomas. En este caso particular el espacio

de sucesos elementales se compone de una cantidad finita de puntos, a los cuales se les asignan probabilidades idénticas. En las etapas iniciales del desarrollo de la teoría de la probabilidad, se consideraban únicamente modelos de este tipo (utilizados para describir juegos de azar, relacionados con el lanzamiento de monedas o dados, con la extracción de una o varias cartas de un mazo de cartas y con situaciones similares), y la igualdad (1.5) servía como definición de probabilidad del suceso \mathbf{A} (es la denominada *definición clásica de probabilidad*).

La violación de la condición de equiprobabilidad de los sucesos elementales (1.4) impide la aplicación de la igualdad (1.5).

Ejemplo 1.4. Calcular la probabilidad de obtener una cantidad impar de puntos al arrojar un dado, por única vez.

Solución. El lanzamiento de un dado tiene 6 resultados posibles, que consideramos equiprobables. (Cuando no se indique lo contrario, consideramos que los dados están equilibrados, y que los 6 resultados posibles de su lanzamiento son equiprobables.) Sea \mathbf{A} el suceso consistente en la aparición de un número impar de puntos. Es evidente que los casos favorables para la ocurrencia del suceso \mathbf{A} son los sucesos consistentes en la aparición de 1, 3 ó 5 puntos. Aplicando la igualdad (1.5), obtenemos $\mathbf{P}(\mathbf{A}) = 3/6 = 1/2$.

Ejemplo 1.5. Calcular la probabilidad de que al tirar un dado dos veces consecutivas, la suma de los puntos obtenidos sea no menor que 8.

Solución. Designemos por (i, j) al resultado del experimento consistente en tirar un dado dos veces consecutivas, y obtener i puntos en el primer tiro y j puntos en el segundo ($i, j = 1, 2, 3, 4, 5, 6$). El conjunto de sucesos elementales que describe los resultados de un experimento de este tipo se compone de $6 \times 6 = 36$ puntos de la forma (i, j) , y puede ser representado en la siguiente tabla:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

El suceso \mathbf{A} consiste en que la suma de los puntos obtenidos es no menor que 8. Es claro que los casos favorables para la ocurrencia del suceso \mathbf{A} son los son indicados en la tabla. La cantidad de estos sucesos es 15.

Considerando que los 36 resultados posibles son equiprobables, y aplicando la fórmula (1.5), obtenemos $\mathbf{P}(\mathbf{A}) = 15/36 = 5/12$.

Ejemplo 1.6. Una urna contiene a bolas blancas y b bolas negras. Se eligen al azar c bolas, donde $c < a + b$. Calcular la probabilidad de que entre las bolas extraídas haya a_0 bolas blancas y b_0 bolas negras.

Solución. Tenemos $a_0 + b_0 = c$, con $0 \leq a_0 \leq a$ y $0 \leq b_0 \leq b$. Es claro que la cantidad de formas distintas de elegir $a_0 + b_0$ bolas entre las $a + b$ bolas de la urna es $C_{a_0 + b_0}^{a + b}$. Todas estas formas serán consideradas equiprobables. Continuando, la cantidad de formas distintas de elegir a_0 bolas blancas entre las a bolas blancas que hay en la urna es $C_{a_0}^a$, y para cada una de las formas de elección de a_0 bolas blancas, existen $C_{b_0}^b$ formas distintas de elegir b_0 bolas negras entre las b bolas negras de la urna. Por ésto, la cantidad de casos favorables para la ocurrencia del suceso \mathbf{A} , consistente en elegir a_0 bolas blancas y b_0 bolas negras, es $C_{a_0}^a C_{b_0}^b$. Según la igualdad (1.5), tenemos

$$\mathbf{P}(\mathbf{A}) = \frac{\binom{a}{a_0} \binom{b}{b_0}}{\binom{a+b}{a_0+b_0}}.$$

1.5. Probabilidad condicional. Fórmulas de la probabilidad total y de Bayes.

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y dos sucesos cualesquiera \mathbf{A}, \mathbf{B} , con $\mathbf{P}(\mathbf{A}) > 0$. Definimos la *probabilidad condicional* de \mathbf{B} dado \mathbf{A} , que designamos $\mathbf{P}(\mathbf{B} | \mathbf{A})$, mediante la fórmula

$$\mathbf{P}(\mathbf{B} | \mathbf{A}) = \frac{\mathbf{P}(\mathbf{AB})}{\mathbf{P}(\mathbf{A})}. \quad (1.6)$$

Veamos que la probabilidad condicional así definida (dado el suceso \mathbf{A} fijo) verifica todos los axiomas de la sección 1.2. Es claro que $\mathbf{P}(\mathbf{B} | \mathbf{A}) \geq 0$ para cualquier suceso \mathbf{B} , de forma que el axioma I se verifica. Continuando,

$$\mathbf{P}(\Omega | \mathbf{A}) = \frac{\mathbf{P}(\Omega \mathbf{A})}{\mathbf{P}(\mathbf{A})} = \frac{\mathbf{P}(\mathbf{A})}{\mathbf{P}(\mathbf{A})} = 1,$$

y el axioma II también se verifica.

Sea $\mathbf{B}_1, \mathbf{B}_2, \dots$ un conjunto finito o numerable de sucesos incompatibles dos a dos. Tenemos

$$\begin{aligned} \mathbf{P}\left(\bigcup_i \mathbf{B}_i \mid \mathbf{A}\right) &= \frac{\mathbf{P}(\mathbf{A}(\bigcup_i \mathbf{B}_i))}{\mathbf{P}(\mathbf{A})} = \frac{\mathbf{P}(\bigcup_i \mathbf{A}\mathbf{B}_i)}{\mathbf{P}(\mathbf{A})} = \frac{\sum_i \mathbf{P}(\mathbf{A}\mathbf{B}_i)}{\mathbf{P}(\mathbf{A})} \\ &= \sum_i \frac{\mathbf{P}(\mathbf{A}\mathbf{B}_i)}{\mathbf{P}(\mathbf{A})} = \sum_i \mathbf{P}(\mathbf{B}_i \mid \mathbf{A}), \end{aligned}$$

y se verifica el axioma III. En conclusión, si $(\Omega, \mathcal{A}, \mathbf{P})$ es un espacio de probabilidad, la terna $(\Omega, \mathcal{A}, \mathbf{P}(\cdot \mid \mathbf{A}))$ donde \mathbf{A} es un suceso con probabilidad positiva, también resulta ser un espacio de probabilidad.

Consideremos ahora el caso particular en el que Ω está compuesto por n puntos, a los cuales se les asignan probabilidades idénticas. De esta forma, es aplicable la regla clásica del cálculo de probabilidades. Para un suceso \mathbf{C} arbitrario, designamos mediante $n_{\mathbf{C}}$ la cantidad de sucesos elementales que componen \mathbf{C} . Entonces $\mathbf{P}(\mathbf{C}) = n_{\mathbf{C}}/n$, y para la probabilidad condicional tenemos

$$\mathbf{P}(\mathbf{B} \mid \mathbf{A}) = \frac{\mathbf{P}(\mathbf{A}\mathbf{B})}{\mathbf{P}(\mathbf{A})} = \frac{n_{\mathbf{A}\mathbf{B}}/n}{n_{\mathbf{A}}/n} = \frac{n_{\mathbf{A}\mathbf{B}}}{n_{\mathbf{A}}}.$$

Ejemplo 1.7. Un experimento consiste en elegir al azar una carta de un mazo de 52 cartas. El suceso \mathbf{A} consiste en que la carta elegida sea roja; el suceso \mathbf{B} , en que sea de corazones. Tenemos $n = 52$, $n_{\mathbf{A}} = 26$, $n_{\mathbf{A}\mathbf{B}} = n_{\mathbf{B}} = 13$, y por ésto

$$\mathbf{P}(\mathbf{B} \mid \mathbf{A}) = \frac{n_{\mathbf{A}\mathbf{B}}}{n_{\mathbf{A}}} = \frac{13}{26} = \frac{1}{2}.$$

Teorema 1.1. *Consideremos sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ incompatibles dos a dos, tales que alguno de ellos ocurre, y con probabilidades positivas. Sea \mathbf{B} un suceso arbitrario. Entonces*

$$\mathbf{P}(\mathbf{B}) = \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) \mathbf{P}(\mathbf{B} \mid \mathbf{A}_i). \quad (1.7)$$

La igualdad (1.7) se denomina fórmula de la probabilidad total.

Demostración. Escribimos $\mathbf{B} = \Omega\mathbf{B} = \bigcup_{i=1}^n \mathbf{A}_i\mathbf{B}$ donde $\mathbf{A}_1\mathbf{B}, \dots, \mathbf{A}_n\mathbf{B}$ son incompatibles dos a dos, por ser dos a dos incompatibles los sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$. Aplicando el axioma III, tenemos

$$\mathbf{P}(\mathbf{B}) = \mathbf{P}\left(\bigcup_{i=1}^n \mathbf{A}_i\mathbf{B}\right) = \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i\mathbf{B}) = \sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) \mathbf{P}(\mathbf{B} \mid \mathbf{A}_i),$$

donde para obtener la última igualdad nos basamos en la definición de probabilidad condicional (1.6). \square

Ejemplo 1.8. Tenemos 5 cajones con productos de una cierta industria. Dos cajones contienen cada uno 4 productos buenos y 1 fallado; otros dos cajones contienen cada uno 3 productos buenos y 2 fallados; y el último cajón contiene 6 productos buenos. Se elige al azar un cajón, del cual, también al azar, se extrae un producto. Calcular la probabilidad de que el producto extraído resulte bueno.

Solución. Designemos mediante \mathbf{B} al suceso consistente en que el producto extraído sea bueno. Tenemos cajones con tres composiciones distintas de productos, y designamos mediante \mathbf{A}_i ($i = 1, 2, 3$) al suceso consistente en elegir un cajón con una de las composiciones dada. De esta forma (por ejemplo) se tiene: el suceso \mathbf{A}_1 consiste en elegir un cajón conteniendo 4 productos buenos y 1 fallado; el suceso \mathbf{A}_2 consiste en elegir un cajón conteniendo 3 productos buenos y 2 fallados; el suceso \mathbf{A}_3 consiste en elegir el cajón que contiene 6 productos buenos. Es claro que los sucesos \mathbf{A}_1 , \mathbf{A}_2 y \mathbf{A}_3 son incompatibles dos a dos y alguno de ellos ocurre, de modo que según la fórmula (1.7), tenemos

$$\begin{aligned} \mathbf{P}(\mathbf{B}) &= \mathbf{P}(\mathbf{A}_1) \mathbf{P}(\mathbf{B} | \mathbf{A}_1) + \mathbf{P}(\mathbf{A}_2) \mathbf{P}(\mathbf{B} | \mathbf{A}_2) + \mathbf{P}(\mathbf{A}_3) \mathbf{P}(\mathbf{B} | \mathbf{A}_3) \\ &= \frac{2}{5} \times \frac{4}{5} + \frac{2}{5} \times \frac{3}{5} + \frac{1}{5} \times \frac{6}{6} = \frac{19}{25}. \end{aligned}$$

Ejemplo 1.9. En una cierta población de hombres hay un 30% de fumadores. Se sabe que la probabilidad de enfermarse de cáncer de pulmón es igual a 0,1 para los fumadores, e igual a 0,01 para los no fumadores. Calcular la probabilidad de que un hombre elegido al azar en esta población esté enfermo de cáncer de pulmón.

Solución. Designemos con la letra \mathbf{B} al suceso consistente en que el hombre elegido tenga esta enfermedad. El suceso \mathbf{A} consiste en elegir un fumador de la población. Sabemos que $\mathbf{P}(\mathbf{A}) = 0,3$, y que $\mathbf{P}(\bar{\mathbf{A}}) = 0,7$ (el suceso $\bar{\mathbf{A}}$ consiste en elegir un no fumador de la población). Por la fórmula de la probabilidad total, tenemos

$$\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{A}) \mathbf{P}(\mathbf{B} | \mathbf{A}) + \mathbf{P}(\bar{\mathbf{A}}) \mathbf{P}(\mathbf{B} | \bar{\mathbf{A}}) = 0,3 \times 0,1 + 0,7 \times 0,01 = 0,037.$$

Teorema 1.2. Consideremos sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ incompatibles dos a dos, tales que alguno de ellos ocurre, y con probabilidades positivas. Sea \mathbf{B} un

suceso con probabilidad positiva. Entonces

$$\mathbf{P}(\mathbf{A}_k | \mathbf{B}) = \frac{\mathbf{P}(\mathbf{A}_k) \mathbf{P}(\mathbf{B} | \mathbf{A}_k)}{\sum_{i=1}^n \mathbf{P}(\mathbf{A}_i) \mathbf{P}(\mathbf{B} | \mathbf{A}_i)} \quad (k = 1, \dots, n). \quad (1.8)$$

La igualdad (1.8) se denomina fórmula de Bayes.

Demostración. Por la definición de probabilidad condicional, tenemos

$$\mathbf{P}(\mathbf{A}_k \mathbf{B}) = \mathbf{P}(\mathbf{A}_k) \mathbf{P}(\mathbf{B} | \mathbf{A}_k) = \mathbf{P}(\mathbf{B}) \mathbf{P}(\mathbf{A}_k | \mathbf{B}) \quad (k = 1, \dots, n).$$

De aquí obtenemos

$$\mathbf{P}(\mathbf{A}_k | \mathbf{B}) = \frac{\mathbf{P}(\mathbf{A}_k) \mathbf{P}(\mathbf{B} | \mathbf{A}_k)}{\mathbf{P}(\mathbf{B})} \quad (k = 1, \dots, n).$$

Para obtener (1.8), resta aplicar la fórmula de la probabilidad total (1.7) en el denominador. \square

Ejemplo 1.10. En una primera urna se tienen 9 bolas blancas y 1 negra, en una segunda urna 2 bolas blancas y 8 negras. Se elige al azar una urna, y de ella, también al azar, se extrae una bola. La bola extraída resulta ser blanca (ocurrió el suceso \mathbf{B}). Calcular las probabilidades $\mathbf{P}(\mathbf{A}_1 | \mathbf{B})$ y $\mathbf{P}(\mathbf{A}_2 | \mathbf{B})$, donde el suceso \mathbf{A}_i consiste en elegir la urna i ($i = 1$ ó 2).

Solución. Por la fórmula de Bayes, tenemos

$$\begin{aligned} \mathbf{P}(\mathbf{A}_1 | \mathbf{B}) &= \frac{\mathbf{P}(\mathbf{A}_1) \mathbf{P}(\mathbf{B} | \mathbf{A}_1)}{\mathbf{P}(\mathbf{A}_1) \mathbf{P}(\mathbf{B} | \mathbf{A}_1) + \mathbf{P}(\mathbf{A}_2) \mathbf{P}(\mathbf{B} | \mathbf{A}_2)} \\ &= \frac{(1/2)(9/10)}{(1/2)(9/10) + (1/2)(2/10)} = \frac{9}{11}. \end{aligned}$$

Análogamente, se obtiene

$$\mathbf{P}(\mathbf{A}_2 | \mathbf{B}) = \frac{(1/2)(2/10)}{(1/2)(9/10) + (1/2)(2/10)} = \frac{2}{11}.$$

Alternativamente, una vez obtenida $\mathbf{P}(\mathbf{A}_1 | \mathbf{B})$ podemos calcular directamente $\mathbf{P}(\mathbf{A}_2 | \mathbf{B}) = 1 - \mathbf{P}(\mathbf{A}_1 | \mathbf{B})$.

1.6. Sucesos independientes

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y dos sucesos cualesquiera \mathbf{A}, \mathbf{B} . Decimos que los sucesos \mathbf{A} y \mathbf{B} son *independientes*, cuando se verifica

$$\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A}) \mathbf{P}(\mathbf{B}). \quad (1.9)$$

Ejemplo 1.11. Se tira un dado dos veces consecutivas. El suceso \mathbf{A} consiste en obtener 6 puntos en primer tiro; el suceso \mathbf{B} , en obtener una cantidad impar de puntos en el segundo. Calculemos la probabilidad de estos sucesos. Como en el ejemplo 1.5 (ver página 24), designamos por (i, j) al resultado correspondiente a obtener i puntos en el primer tiro y j puntos en el segundo ($i, j = 1, 2, 3, 4, 5, 6$). Hay $6 \times 6 = 36$ sucesos elementales de esta forma. Los casos favorables para la ocurrencia del suceso \mathbf{A} son los puntos $(6, 1), (6, 2), (6, 3), (6, 4), (6, 5)$ y $(6, 6)$; por la regla clásica del cálculo de probabilidades, tenemos $\mathbf{P}(\mathbf{A}) = 6/36 = 1/6$. Los casos favorables para la ocurrencia del suceso \mathbf{B} son los de la forma $(i, 1), (i, 3), (i, 5)$, en donde $1 \leq i \leq 6$, por lo que $\mathbf{P}(\mathbf{B}) = 18/36 = 1/2$. Los casos favorables para la ocurrencia del suceso \mathbf{AB} son $(6, 1), (6, 3)$ y $(6, 5)$, de forma que $\mathbf{P}(\mathbf{AB}) = 3/36 = 1/12$. Como

$$\mathbf{P}(\mathbf{A}) \mathbf{P}(\mathbf{B}) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12} = \mathbf{P}(\mathbf{AB}),$$

los sucesos \mathbf{A} y \mathbf{B} son independientes.

Ejemplo 1.12. Consideremos el experimento correspondiente a elegir una carta al azar en un mazo de 52 cartas. El suceso \mathbf{A} consiste en que la carta sea una figura; y el suceso \mathbf{B} , en que sea negra. Demostremos que los sucesos \mathbf{A} y \mathbf{B} son independientes. En efecto, por la regla clásica del cálculo de probabilidades, tenemos

$$\mathbf{P}(\mathbf{A}) = \frac{12}{52} = \frac{3}{13}, \quad \mathbf{P}(\mathbf{B}) = \frac{26}{52} = \frac{1}{2}, \quad \mathbf{P}(\mathbf{AB}) = \frac{6}{52} = \frac{3}{26},$$

de forma que se cumple la igualdad (1.9).

Veamos ahora que si ambos sucesos \mathbf{A} y \mathbf{B} tienen probabilidad positiva (esto asegura que $\mathbf{P}(\mathbf{B}|\mathbf{A})$ y $\mathbf{P}(\mathbf{A}|\mathbf{B})$ están definidas), entonces, la independencia de los sucesos \mathbf{A} y \mathbf{B} es equivalente a alguna de las igualdades

$$\mathbf{P}(\mathbf{B}|\mathbf{A}) = \mathbf{P}(\mathbf{B}), \quad (1.10)$$

$$\mathbf{P}(\mathbf{A}|\mathbf{B}) = \mathbf{P}(\mathbf{A}). \quad (1.11)$$

En efecto, si \mathbf{A} y \mathbf{B} son independientes, tenemos $\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})$. Por otra parte, $\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B}|\mathbf{A}) = \mathbf{P}(\mathbf{B})\mathbf{P}(\mathbf{A}|\mathbf{B})$, y por ésto (1.10) y (1.11) son válidas. Si se cumple una de las igualdades (1.10) ó (1.11), por ejemplo (1.10), entonces

$$\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B}|\mathbf{A}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B}),$$

y los sucesos \mathbf{A}, \mathbf{B} son independientes.

Proposición 1.1. *Si \mathbf{A} y \mathbf{B} son sucesos independientes, se cumple: (i) \mathbf{A} y $\bar{\mathbf{B}}$ son independientes; (ii) $\bar{\mathbf{A}}$ y \mathbf{B} son independientes; (iii) $\bar{\mathbf{A}}$ y $\bar{\mathbf{B}}$ son independientes.*

Demostración. Para sucesos \mathbf{A} y \mathbf{B} arbitrarios tenemos $\mathbf{A} = \mathbf{AB} \cup \mathbf{A}\bar{\mathbf{B}}$, por lo que $\mathbf{P}(\mathbf{A}) = \mathbf{P}(\mathbf{AB}) + \mathbf{P}(\mathbf{A}\bar{\mathbf{B}})$. Si \mathbf{A} y \mathbf{B} son independientes, en vista de (1.9) tenemos $\mathbf{P}(\mathbf{A}\bar{\mathbf{B}}) = \mathbf{P}(\mathbf{A}) - \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{A})(1 - \mathbf{P}(\mathbf{B})) = \mathbf{P}(\mathbf{A})\mathbf{P}(\bar{\mathbf{B}})$, lo que demuestra (i). Cambiando los roles entre \mathbf{A} y \mathbf{B} obtenemos (ii). La afirmación (iii) se obtiene aplicando (i) a los sucesos \mathbf{A} y $\bar{\mathbf{B}}$, que son independientes en vista de (ii). \square

Decimos que los sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ son *independientes dos a dos*, cuando se verifica

$$\mathbf{P}(\mathbf{A}_i\mathbf{A}_j) = \mathbf{P}(\mathbf{A}_i)\mathbf{P}(\mathbf{A}_j),$$

para todo $i \neq j$, donde $i, j = 1, \dots, n$.

Decimos que los sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ son *mutuamente independientes*, o mas brevemente, *independientes*, cuando para todo k ($2 \leq k \leq n$) se verifica

$$\mathbf{P}\left(\bigcap_{m=1}^k \mathbf{A}_{i_m}\right) = \prod_{m=1}^k \mathbf{P}(\mathbf{A}_{i_m}), \quad (1.12)$$

para cualquier elección de naturales i_1, \dots, i_k , que cumplan la condición $1 \leq i_1 < i_2 < \dots < i_k \leq n$.

De esta forma, la independencia mutua de tres sucesos \mathbf{A}, \mathbf{B} y \mathbf{C} , significa que se cumplen las igualdades $\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})$, $\mathbf{P}(\mathbf{AC}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{C})$ y $\mathbf{P}(\mathbf{BC}) = \mathbf{P}(\mathbf{B})\mathbf{P}(\mathbf{C})$ (que implican la independencia dos a dos de los sucesos \mathbf{A}, \mathbf{B} y \mathbf{C}), y también $\mathbf{P}(\mathbf{ABC}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})\mathbf{P}(\mathbf{C})$.

Veamos que la independencia dos a dos de n sucesos ($n \geq 3$), en general, no implica la independencia mutua de estos sucesos. Con este fin, consideramos el siguiente ejemplo, propuesto por S. N. Bernstein.

Ejemplo 1.13. Sea Ω un espacio de sucesos elementales compuesto por los puntos $\omega_1, \omega_2, \omega_3$ y ω_4 . Consideremos los sucesos $\mathbf{A} = \{\omega_1, \omega_4\}$, $\mathbf{B} = \{\omega_2, \omega_4\}$, y $\mathbf{C} = \{\omega_3, \omega_4\}$. Asignemos las probabilidades

$$\mathbf{P}(\omega_1) = \mathbf{P}(\omega_2) = \mathbf{P}(\omega_3) = \mathbf{P}(\omega_4) = \frac{1}{4}.$$

Entonces $\mathbf{P}(\mathbf{A}) = 2/4 = 1/2$. Análogamente obtenemos $\mathbf{P}(\mathbf{B}) = \mathbf{P}(\mathbf{C}) = 1/2$. Además, $\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{AC}) = \mathbf{P}(\mathbf{BC}) = \mathbf{P}(\omega_4) = 1/4$. De esta forma se verifican las tres igualdades $\mathbf{P}(\mathbf{AB}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})$, $\mathbf{P}(\mathbf{AC}) = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{C})$, y $\mathbf{P}(\mathbf{BC}) = \mathbf{P}(\mathbf{B})\mathbf{P}(\mathbf{C})$, y en consecuencia, los sucesos \mathbf{A} , \mathbf{B} y \mathbf{C} son independientes dos a dos. Sin embargo,

$$\mathbf{P}(\mathbf{ABC}) = \mathbf{P}(\omega_4) = \frac{1}{4} \neq \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})\mathbf{P}(\mathbf{C}) = \frac{1}{8},$$

por lo que nos sucesos \mathbf{A} , \mathbf{B} y \mathbf{C} no son mutuamente independientes.

El ejemplo de Bernstein se puede formular de otra manera. Consideremos el experimento consistente en arrojar un tetraedro, cuyas caras están coloreadas de la siguiente forma: una cara es roja; otra cara es azul; una tercer cara es verde; y la cuarta cara tiene los tres colores indicados. El suceso \mathbf{A} consiste en la presencia del color rojo en la cara sobre la que se apoya el tetraedro al caer, el suceso \mathbf{B} consiste en la presencia del azul, y el \mathbf{C} , en la presencia del verde. Los sucesos \mathbf{A} , \mathbf{B} y \mathbf{C} son independientes dos a dos, pero no son mutuamente independientes.

1.7. Ejercicios

1. Un blanco se compone de 5 círculos concéntricos con radios $0 < r_1 < r_2 < r_3 < r_4 < r_5$. El suceso \mathbf{A}_k consiste en acertar en el círculo de radio r_k . Explicar que significan los sucesos $\mathbf{B} = \bigcup_{k=1}^5 \mathbf{A}_k$, $\mathbf{C} = \bigcap_{k=1}^5 \mathbf{A}_k$, y $\mathbf{D} = \bar{\mathbf{A}}_1 \mathbf{A}_2$.
2. Demostrar que para dos sucesos \mathbf{A} y \mathbf{B} arbitrarios, las siguientes cuatro relaciones son equivalentes: (a) $\mathbf{A} \subset \mathbf{B}$; (b) $\bar{\mathbf{B}} \subset \bar{\mathbf{A}}$; (c) $\mathbf{A} \cup \mathbf{B} = \mathbf{B}$; (d) $\mathbf{A}\bar{\mathbf{B}} = \emptyset$.
3. Un trabajador fabrica distintos productos. Sea \mathbf{A}_k ($k = 1, \dots, n$) el suceso que consiste en que el producto k -ésimo sea defectuoso. Escribir los sucesos: (a) ni uno de los productos es defectuoso; (b) por lo menos

uno de los productos es defectuoso; (c) solamente uno de los productos es defectuoso.

4. Se tiran dos dados en forma consecutiva. El suceso \mathbf{A} consiste en que la suma de puntos obtenidos sea par; el suceso \mathbf{B} , en que por lo menos en uno de los dados aparezcan 6 puntos. Describa los sucesos $\mathbf{A} \cup \mathbf{B}$, $\mathbf{A}\mathbf{B}$, $\mathbf{A}\bar{\mathbf{B}}$, $\bar{\mathbf{A}}\mathbf{B}$.

5. Sean \mathbf{A} y \mathbf{B} sucesos arbitrarios. El suceso $(\mathbf{A} \setminus \mathbf{B}) \cup (\mathbf{B} \setminus \mathbf{A})$ se denomina *diferencia simétrica* entre los sucesos \mathbf{A} y \mathbf{B} , y se designa mediante $\mathbf{A} \triangle \mathbf{B}$. Demostrar que: (a) $\mathbf{A} \cup \mathbf{B} = (\mathbf{A}\mathbf{B}) \cup (\mathbf{A} \triangle \mathbf{B})$; (b) $\mathbf{A} \triangle \bar{\mathbf{A}} = \Omega$; (c) $\mathbf{A} \triangle \Omega = \bar{\mathbf{A}}$.

6. Demostrar que para cualquier sucesión de sucesos $\mathbf{A}_1, \mathbf{A}_2, \dots$ vale la igualdad

$$\bigcup_{n=1}^{\infty} \mathbf{A}_n = \mathbf{A}_1 \cup (\bar{\mathbf{A}}_1 \mathbf{A}_2) \cup (\bar{\mathbf{A}}_1 \bar{\mathbf{A}}_2 \mathbf{A}_3) \cup \dots$$

7. Demostrar que si $\mathbf{A}_1 \subset \mathbf{A}_2 \subset \mathbf{A}_3 \subset \dots$, entonces existe el límite $\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_n) = \mathbf{P}(\mathbf{A})$, donde $\mathbf{A} = \bigcup_{n=1}^{\infty} \mathbf{A}_n$.

8. Demostrar que $\mathbf{P}(\mathbf{A}\mathbf{B}) \geq \mathbf{P}(\mathbf{A}) - \mathbf{P}(\bar{\mathbf{B}})$ para sucesos \mathbf{A} y \mathbf{B} arbitrarios.

9. Demostrar que se verifica $\mathbf{P}(\bigcup_{k=1}^n \mathbf{A}_k) \geq 1 - \sum_{k=1}^n \mathbf{P}(\bar{\mathbf{A}}_k)$ para sucesos $\mathbf{A}_1, \dots, \mathbf{A}_k$ arbitrarios.

10. Demostrar que $\mathbf{P}(\bigcup_{k=1}^n \mathbf{A}_k) = 1 - \mathbf{P}(\bigcap_{k=1}^n \bar{\mathbf{A}}_k) \leq \sum_{k=1}^n \mathbf{P}(\mathbf{A}_k)$, para sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ arbitrarios.

11. Demostrar que $\mathbf{P}(\mathbf{A} \setminus \mathbf{B}) = \mathbf{P}(\mathbf{A}) - \mathbf{P}(\mathbf{A}\mathbf{B})$ para sucesos \mathbf{A} y \mathbf{B} arbitrarios.

12. Una urna contiene 4 bolas blancas y 5 negras. Se eligen tres bolas al azar. Calcular las probabilidades de que: (a) todas las bolas extraídas sean blancas; (b) todas las bolas extraídas sean negras; (c) se extraiga una bola blanca y dos negras.

13. Para obtener el premio mayor en una lotería se precisa acertar 5 números elegidos entre 49. Calcular la probabilidad de obtener el premio mayor en esta lotería.

14. De un mazo de 52 cartas se eligen 4 cartas al azar. Calcular la probabilidad de que se extraigan: (a) por lo menos un as; (b) no menos de dos ases.

15. Se considera un experimento consistente en arrojar un dado dos veces consecutivas. Calcular la probabilidad de que la suma de los resultados sea: (a) igual a 5; (b) no mayor de 5.

16. Hallar la probabilidad de que al tirar un dado tres veces consecutivas, la suma de los resultados sea no menor que 16.

17. Calcular la probabilidad de que se acepte una partida de 100 unidades, 5 de las cuales están falladas, si se toman de muestra la mitad, y las condiciones para aceptarla son contener a lo sumo un 2% de unidades falladas.

18. En el ejercicio 4 calcular las probabilidades de los sucesos $\mathbf{A} \cup \mathbf{B}$, $\mathbf{A}\mathbf{B}$, $\mathbf{A}\bar{\mathbf{B}}$, $\bar{\mathbf{A}}\mathbf{B}$.

19. Demostrar la igualdad

$$\begin{aligned} \mathbf{P}\left(\bigcup_{n=1}^{\infty} \mathbf{A}_n\right) &= \mathbf{P}(\mathbf{A}_1) + \mathbf{P}(\bar{\mathbf{A}}_1\mathbf{A}_2) + \mathbf{P}(\bar{\mathbf{A}}_1\bar{\mathbf{A}}_2\mathbf{A}_3) + \cdots \\ &\quad + \mathbf{P}(\bar{\mathbf{A}}_1 \dots \bar{\mathbf{A}}_{n-1}\mathbf{A}_n) + \cdots \end{aligned}$$

donde $\mathbf{A}_1, \mathbf{A}_2 \dots$ es una sucesión de sucesos arbitrarios.

20. Se tienen K urnas con n bolas cada una, numeradas de 1 a n . De cada urna se elige al azar una bola. Hallar la probabilidad de que el número mayor resultante sea m ($m = 1, \dots, n$).

21. Tres jugadores A, B y C extraen por turno una bola cada uno, de una urna que contiene 10 bolas blancas y 10 bolas negras. Las bolas extraídas no se reponen, y gana el primero que extrae una bola blanca. Calcular la probabilidad de que gane cada uno de los jugadores A, B, y C.

22. Sean \mathbf{A} y \mathbf{B} dos sucesos arbitrarios, con $\mathbf{P}(\mathbf{A}) > 0$. Demostrar la desigualdad

$$\mathbf{P}(\mathbf{B}|\mathbf{A}) \geq 1 - \frac{\mathbf{P}(\bar{\mathbf{B}})}{\mathbf{P}(\mathbf{A})}.$$

- 23.** De una urna que contiene 4 bolas blancas y 2 negras se extrae al azar una bola, que luego se pone en una segunda urna, que tiene 3 bolas blancas y 4 negras. Calcular la probabilidad de que una bola extraída de la segunda urna sea blanca.
- 24.** Un estudiante asiste a un examen sabiendo solo 15 de las 20 preguntas del programa. En el billete del examen hay 3 preguntas. Calcular la probabilidad de que el estudiante sepa las 3 preguntas, de las dos formas siguientes: (a) aplicando las reglas clásicas del cálculo de probabilidades; (b) utilizando la noción de probabilidad condicional.
- 25.** En una mesa hay tres armas de tipo A y una de tipo B. La probabilidad de acertar en el blanco con un arma de tipo A es de 0,7, y la de acertar con un arma de tipo B es 0,4. Se elige al azar un arma y se dispara un tiro al blanco. Calcular: (a) la probabilidad de fallar el tiro; (b) la probabilidad de haber elegido un arma de tipo B, sabiendo que el tiro falló.
- 26.** En una caja hay 4 pelotas de tenis nuevas y 2 usadas. Para un primer partido, se eligen 2 pelotas al azar, y luego se retornan a la caja. Se eligen otras dos pelotas de la misma caja para un segundo partido. Calcular la probabilidad de que ambas sean nuevas.
- 27.** Los sucesos **A**, **B** y **C** son tales que: **A** y **B** son independientes; **A** y **C** son incompatibles; **B** y **C** son independientes; $P(\mathbf{A}) = 0,6$, $P(\mathbf{B}) = 0,4$ y $P(\mathbf{C}) = 0,1$. Calcular las probabilidades de los sucesos $\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}$ y $\mathbf{A}\bar{\mathbf{B}}$.
- 28.** Demostrar que si los sucesos **A** y **B** son independientes, y ambos tienen probabilidad positiva, entonces no pueden ser incompatibles.
- 29.** Demostrar que si **A** es un suceso arbitrario, y **B** es tal que $P(\mathbf{B}) = 0$, entonces **A** y **B** son independientes.
- 30.** Sean **A** y **B** dos sucesos independientes, y tales que $\mathbf{A} \subset \mathbf{B}$. Demostrar que si $P(\mathbf{A}) \neq 0$, entonces $P(\mathbf{B}) = 1$.
- 31.** La probabilidad de detectar un avión que vuela en una determinada región, por medio de un radar, es 0,9. En esta región operan en forma independiente tres radares. Calcular la probabilidad de que se detecte un avión en esa zona: (a) mediante los tres radares; (b) mediante por lo menos un radar.

32. En la fabricación de un cierto aparato se utilizan dos piezas del mismo tipo. Para que el aparato funcione, se precisa que por lo menos una de las piezas no esté fallada. La probabilidad de que la pieza esté fallada es 0,05. Calcular, bajo el supuesto de independencia, la probabilidad de que el mencionado aparato funcione.

33. Sean **A**, **B** y **C** sucesos independientes dos a dos y equiprobables, cada uno de los cuales tiene probabilidad p . Supongamos que $\mathbf{P}(\mathbf{ABC}) = 0$. Hallar el valor de p que hace que la probabilidad de el suceso $\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}$ sea máxima.

Capítulo 2

Esquema de Bernoulli

2.1. Esquema de Bernoulli y fórmula de la distribución binomial

Consideremos un espacio de sucesos elementales Ω compuesto por dos puntos ω_1 y ω_2 . Llamamos *éxito* al punto ω_1 y *fracaso* al punto ω_2 . Sea \mathcal{A} el conjunto de todos los subconjuntos del espacio Ω . Introducimos una probabilidad \mathbf{P} asignándole a los puntos ω_1 y ω_2 , como probabilidades, dos números positivos p y q que verifican $p + q = 1$ (ver ejemplo 1.1).

Consideremos ahora un nuevo espacio de sucesos elementales Ω_2 compuesto por los cuatro puntos (ω_1, ω_1) , (ω_1, ω_2) , (ω_2, ω_1) y (ω_2, ω_2) . Este espacio describe el conjunto de todos los resultados posibles de dos experimentos, cada uno de los cuales puede concluir solamente con uno de dos resultados: éxito ω_1 , o fracaso ω_2 . Por ejemplo, el punto (ω_1, ω_2) corresponde a la ocurrencia de un éxito en el primer experimento, y un fracaso en el segundo. Sea \mathcal{A}_2 el conjunto de todos los subconjuntos del espacio Ω_2 . Introducimos una probabilidad \mathbf{P} asignándole a cada uno de los cuatro puntos las probabilidades siguientes: p^2 , pq , pq y q^2 (ver ejemplo 1.2). La suma de estas cuatro probabilidades es $p^2 + 2pq + q^2 = (p + q)^2 = 1$. El espacio de probabilidad $(\Omega_2, \mathcal{A}_2, \mathbf{P})$, donde \mathbf{P} es la probabilidad recién definida, se denomina *serie de dos experimentos independientes*.

Consideremos una serie de dos experimentos independientes. El suceso \mathbf{A} consiste en la ocurrencia de un éxito en el primer experimento; el \mathbf{B} , en la ocurrencia de un éxito en el segundo. De esta forma

$$\mathbf{A} = \{(\omega_1, \omega_1), (\omega_1, \omega_2)\}, \quad \mathbf{B} = \{(\omega_1, \omega_1), (\omega_2, \omega_1)\}.$$

Veamos que los sucesos \mathbf{A} y \mathbf{B} son independientes. En efecto, tenemos $\mathbf{P}(\mathbf{A}) = p^2 + pq = p(p+q) = p$, y análogamente, $\mathbf{P}(\mathbf{B}) = p^2 + pq = p$. Como $\mathbf{AB} = \{(\omega_1, \omega_1)\}$, resulta que $\mathbf{P}(\mathbf{AB}) = p^2 = \mathbf{P}(\mathbf{A})\mathbf{P}(\mathbf{B})$, verificándose la definición de sucesos independientes (ver sección 1.6). De acuerdo a la proposición 1.1, también son independientes los sucesos \mathbf{A} y $\bar{\mathbf{B}}$, $\bar{\mathbf{A}}$ y \mathbf{B} , $\bar{\mathbf{A}}$ y $\bar{\mathbf{B}}$. De esta forma, en una serie de dos experimentos independientes, los resultados correspondientes a cada uno de los experimentos son sucesos independientes.

Consideremos, más en general, un espacio de sucesos elementales Ω_n , compuesto por los puntos de la forma

$$\omega = (\omega^{(1)}, \dots, \omega^{(n)}), \quad (2.1)$$

donde cada $\omega^{(i)}$ es o bien éxito ω_1 , o bien fracaso ω_2 ($i = 1, \dots, n$). El espacio Ω_n describe el conjunto de todos los resultados posibles de n experimentos, cada uno de los cuales puede concluir solamente con uno de dos resultados: éxito ω_1 , o fracaso ω_2 . Sea \mathcal{A}_n el conjunto de todos los subconjuntos del espacio Ω_n . Dado un suceso de la σ -álgebra \mathcal{A}_n , le asignamos una probabilidad igual a la suma de las probabilidades de los sucesos elementales que lo componen; mientras que la probabilidad de cada suceso elemental del espacio Ω_n se define de la siguiente forma: si el punto $(\omega^{(1)}, \dots, \omega^{(n)})$ tiene m componentes ω_1 y $n - m$ componentes ω_2 , le asignamos una probabilidad $p^m q^{n-m}$. (Luego veremos que la suma de las probabilidades de todos los sucesos elementales es igual a 1, lo que permite verificar los tres axiomas introducidos en la sección 1.2.)

El espacio de probabilidad $(\Omega_n, \mathcal{A}_n, \mathbf{P})$, donde \mathbf{P} es la probabilidad recién definida, se denomina *serie de n experimentos independientes*, o *esquema de Bernoulli*.

Designemos mediante $\mu(\omega)$ la cantidad de componentes iguales a ω_1 en el suceso elemental ω dado en (2.1). De esta forma, $\mu(\omega)$ es la cantidad de éxitos en n experimentos. Introduzcamos la notación

$$P_n(m) = \mathbf{P}(\{\omega : \mu(\omega) = m\}) = \mathbf{P}(\mu = m),$$

para $m = 0, 1, \dots, n$. En palabras, $P_n(m)$ es la probabilidad de que ocurran m éxitos en n experimentos independientes.

Proposición 2.1. *Tiene lugar la igualdad*

$$P_n(m) = \binom{n}{m} p^m q^{n-m}, \quad (m = 0, 1, \dots, n). \quad (2.2)$$

Demostración. Queremos calcular la probabilidad de que, al realizar n experimentos, ocurran exactamente m éxitos. De acuerdo con nuestra definición, dicha probabilidad es la suma de las probabilidades de los puntos ω de la forma dada en (2.1), que componen el suceso $\{\omega: \mu(\omega) = m\}$. Un punto perteneciente a este suceso es, por ejemplo,

$$\omega = (\underbrace{\omega_1, \dots, \omega_1}_m, \underbrace{\omega_2, \dots, \omega_2}_{n-m}),$$

que corresponde a la ocurrencia de éxitos en los m primeros experimentos, y fracasos en los $n - m$ restantes. Este punto tiene una probabilidad igual a $p^m q^{n-m}$. Más aún, todos los puntos con exactamente m éxitos tienen asignada esta misma probabilidad, dado que las probabilidades de los sucesos elementales no dependen del lugar en la serie que ocupan los experimentos en que ocurren los éxitos, sino solamente, de la cantidad total de éxitos. Resta entonces saber, cuantos son los puntos que componen el suceso $\{\omega: \mu(\omega) = m\}$. Es claro que esta cantidad es igual a la cantidad de formas de distribuir m objetos (los éxitos ω_1) en n lugares (las n componentes del punto ω), siendo por ésto igual a $\binom{n}{m}$. En conclusión, se obtiene $P_n(m) = \binom{n}{m} p^m q^{n-m}$, concluyendo la demostración. \square

Observación. Tenemos

$$\sum_{m=0}^n P_n(m) = \sum_{m=0}^n \binom{n}{m} p^m q^{n-m} = (q + p)^n = 1.$$

Esta igualdad muestra que la asignación de probabilidades es correcta (es decir, se verifican los axiomas de la sección 1.2).

Observación. Decimos que la fórmula (2.2) es la *distribución de probabilidades binomiales*.

Estudiemos algunas consecuencias sencillas de la proposición recién demostrada. La probabilidad de que ocurran n éxitos en n experimentos independientes es igual a p^n ; la de que ocurran n fracasos, igual a q^n . (Estos resultados se obtienen de la fórmula (2.2), en los casos $m = n$ y $m = 0$ respectivamente.) La probabilidad de que ocurra por lo menos un éxito en n experimentos independientes es $1 - q^n$, como resulta de aplicar la igualdad $\mathbf{P}(\bar{\mathbf{A}}) = 1 - \mathbf{P}(\mathbf{A})$ (Propiedad 1, sección 1.3).

Consideremos algunos ejemplos.

Ejemplo 2.1. Calcular la probabilidad de que en 5 tiradas consecutivas de una moneda, aparezca cara: (a) 5 veces; (b) ni una vez; (c) por lo menos una vez.

Solución. Considerando las 5 tiradas consecutivas de una moneda como una serie de 5 experimentos independientes, con dos resultados posibles cada uno (cara y número), y suponiendo que la probabilidad de obtener una cara (probabilidad de éxito) es igual a $1/2$, obtenemos: (a) $P_5(5) = (1/2)^5 = 1/32$; (b) $P_5(0) = (1 - 1/2)^5 = 1/32$. (c) Por lo anterior, la probabilidad de que aparezca cara por lo menos una vez, es $1 - 1/32 = 31/32$.

Ejemplo 2.2. Se realiza una serie de tres disparos, con una probabilidad de acertar en el blanco igual a $1/3$. Calcular la probabilidad de que: (a) dos veces se acierte en el blanco; (b) por lo menos una vez se acierte en el blanco.

Solución. Por la fórmula (2.2), con $p = 1/3$, tenemos:

$$(a) P_3(2) = \binom{3}{2} (1/3)^2 (1 - 1/3)^{3-2} = 2/9.$$

$$(b) 1 - P_3(0) = 1 - (1 - 1/3)^3 = 1 - 8/27 = 19/27.$$

Ejemplo 2.3. En determinadas condiciones de producción, la probabilidad de que un cierto artículo resulte defectuoso es igual a $0,02$. Calcular la probabilidad de que en 10000 artículos elegidos al azar, resulten: (a) 230 defectuosos; (b) a lo sumo 230 defectuosos.

Solución. Considerando el control de calidad de 10000 artículos elegidos al azar como una serie de 10000 experimentos independientes, con probabilidad de éxito $p = 0,02$ (llamamos éxito al suceso consistente en que el artículo resulte defectuoso), obtenemos los siguientes resultados:

$$(a) P_{10000}(230) = \binom{10000}{230} (0,02)^{230} (0,98)^{9770},$$

(b) Queremos calcular la probabilidad del suceso $\{\omega: \mu(\omega) \leq 230\}$. Para ésto, sumamos las probabilidades de los sucesos (incompatibles dos a dos) de la forma $\{\omega: \mu(\omega) = m\}$ ($m = 0, \dots, 230$). Entonces

$$\mathbf{P}(\mu \leq 230) = \sum_{m=0}^{230} \binom{10000}{m} (0,02)^m (0,98)^{10000-m}.$$

Las expresiones que hemos obtenido en el ejercicio anterior, sin lugar a duda, son difíciles de calcular numéricamente. Esto muestra la importancia de conocer fórmulas que aproximen a estas cantidades, y permiten hacer

los cálculos hasta el final. Estas fórmulas aproximadas son las proporcionadas por los teoremas límites de De Moivre–Laplace, que estudiaremos en las dos siguientes secciones.

2.2. Teorema límite local de De Moivre–Laplace

Teorema 2.1 (Teorema límite local de De Moivre–Laplace).

Consideremos una serie de n experimentos independientes, con probabilidad de éxito en cada experimento igual a p ($0 < p < 1$), $q = 1 - p$. Sean $P_n(m)$ la probabilidad de obtener m éxitos en n experimentos, y $x = x_{n,m} = \frac{m-np}{\sqrt{npq}}$. Entonces, tiene lugar la convergencia

$$\frac{\sqrt{npq}P_n(m)}{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}} \rightarrow 1, \quad \text{si } n \rightarrow \infty, \quad (2.3)$$

uniformemente en el conjunto de los valores de m tales que $|x_{n,m}| \leq C$, donde C es una constante arbitraria.

Una manera alternativa de escribir la convergencia que tiene lugar en (2.3), es

$$\sup \left| \frac{\sqrt{npq}P_n(m)}{e^{-x^2/2}/\sqrt{2\pi}} - 1 \right| \rightarrow 0, \quad \text{si } n \rightarrow \infty,$$

donde el supremo se toma en el conjunto de valores de m tales que $|x_{n,m}| \leq C$.

Demostración. La demostración se basa en la proposición 2.1 y en la fórmula de Stirling.

En vista de la definición de x , tenemos

$$m = np + x\sqrt{npq}, \quad (2.4)$$

$$n - m = nq - x\sqrt{npq}. \quad (2.5)$$

Estas fórmulas, y la condición $|x_{n,m}| \leq C$, implican que $m \rightarrow \infty$ y $n - m \rightarrow \infty$ cuando $n \rightarrow \infty$. De la fórmula (2.2) se obtiene, que

$$P_n(m) = \frac{n!}{m!(n-m)!} p^m q^{n-m}.$$

Sustituyendo en esta expresión, mediante la *fórmula de Stirling*, que establece la igualdad $n! = n^n e^{-n} \sqrt{2\pi n} (1 + \alpha_n)$, donde $\alpha_n \rightarrow 0$ ($n \rightarrow \infty$), obtenemos,

$$P_n(m) = \frac{n^n e^{-n} \sqrt{2\pi n} (1 + \alpha_n)}{m^m e^{-m} \sqrt{2\pi m} (1 + \alpha_m)} \times \frac{1}{(n-m)^{n-m} e^{-n+m} \sqrt{2\pi(n-m)} (1 + \alpha_{n-m})} \times p^m q^{n-m},$$

que simplificando, y agrupando las potencias de m y de $n-m$, y los términos con α_n , es

$$P_n(m) = \left(\frac{np}{m}\right)^{m+\frac{1}{2}} \left(\frac{nq}{n-m}\right)^{n-m+\frac{1}{2}} \frac{1}{\sqrt{2\pi npq}} (1 + \beta_{n,m}),$$

donde

$$\beta_{n,m} = \frac{1 + \alpha_n}{(1 + \alpha_m)(1 + \alpha_{n-m})} - 1.$$

Observemos, que $\beta_{n,m} \rightarrow 0$, cuando $n \rightarrow \infty$, uniformemente, en el conjunto de los valores de m tales que $|x_{n,m}| \leq C$.

Es conveniente reescribir la fórmula anterior, como

$$\sqrt{2\pi npq} P_n(m) = \left(\frac{m}{np}\right)^{-(m+1/2)} \left(\frac{n-m}{nq}\right)^{-(n-m+1/2)} (1 + \beta_{n,m}).$$

De las expresiones (2.4) y (2.5) se obtiene, que $\frac{m}{np} = 1 + x \sqrt{\frac{q}{np}}$ y $\frac{n-m}{nq} = 1 - x \sqrt{\frac{p}{nq}}$. Esto permite sustituir en los dos primeros factores, para obtener la fórmula

$$\sqrt{2\pi npq} P_n(m) = T_{n,m} (1 + \beta_{n,m}), \quad (2.6)$$

donde el término $T_{n,m}$ está dado por

$$T_{n,m} = \left(1 + x \sqrt{\frac{q}{np}}\right)^{-(m+1/2)} \left(1 - x \sqrt{\frac{p}{nq}}\right)^{-(n-m+1/2)}.$$

Por último, tomando logaritmos naturales (en la base e), se tiene

$$\ln T_{n,m} = -(m+1/2) \ln \left(1 + x \sqrt{\frac{q}{np}}\right) - (n-m+1/2) \ln \left(1 - x \sqrt{\frac{p}{nq}}\right). \quad (2.7)$$

El resto de la demostración consiste en obtener la fórmula $\ln T_{n,m} = -x^2/2 + r_{n,m}$ (equivalente a $T_{n,m} = e^{-x^2/2} e^{r_{n,m}}$), donde $r_{n,m} \rightarrow 0$ ($n \rightarrow \infty$) uniformemente.

Utilizamos entonces el desarrollo de Taylor de la función $\ln(1+u)$, en el punto $u = 0$, que establece que $\ln(1+u) = u - \frac{u^2}{2} + \theta u^3$, si $|u| < 1/2$.

Poniendo $u = x\sqrt{\frac{q}{np}}$ en el primer sumando en (2.7), y $u = -x\sqrt{\frac{p}{nq}}$ en el segundo, y teniendo en cuenta (2.4) y (2.5), resulta

$$\begin{aligned} \ln T_{n,m} = & -(np + x\sqrt{npq} + 1/2) \left(x\sqrt{\frac{q}{np}} - \frac{x^2q}{2np} + \theta_1 \left(x\sqrt{\frac{q}{np}} \right)^3 \right) \\ & - (nq - x\sqrt{npq} + 1/2) \left(-x\sqrt{\frac{p}{nq}} - \frac{x^2p}{2nq} + \theta_2 \left(x\sqrt{\frac{p}{nq}} \right)^3 \right) \end{aligned}$$

donde $|\theta_1| < 3$, $|\theta_2| < 3$, para todo n suficientemente grande. En consecuencia, multiplicando y simplificando, obtenemos

$$\ln T_{n,m} = -x\sqrt{npq} - x^2q + \frac{x^2q}{2} + x\sqrt{npq} - x^2p + \frac{x^2p}{2} + r_{n,m} = -\frac{x^2}{2} + r_{n,m},$$

donde $|r_{n,m}| \leq C_0/\sqrt{n}$, con C_0 una constante que depende únicamente de p, q y C . Sustituyendo en (2.6), obtenemos

$$\sqrt{2\pi npq} P_n(m) = e^{-x^2/2} e^{r_{n,m}} (1 + \beta_{n,m}),$$

que escrita en forma similar al enunciado del teorema, es

$$\frac{\sqrt{npq} P_n(m)}{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}} = e^{r_{n,m}} (1 + \beta_{n,m}),$$

Como $\beta_{n,m}$ y $r_{n,m}$ convergen uniformemente a cero, en el conjunto de los valores de m tales que $|x_{m,n}| \leq C$, la demostración está terminada. \square

En vista del teorema recién demostrado, se obtiene, que para n suficientemente grande, tiene lugar la identidad aproximada

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{m - np}{\sqrt{npq}}\right),$$

donde

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

La función $\varphi(x)$ se denomina *densidad de la distribución normal*. En el final del libro se presenta una tabla con los valores de esta función.

Con esta aproximación, estamos ahora en condiciones de calcular la probabilidad $P_{10000}(230)$ de la parte (a) del ejemplo 2.3. En este caso, tenemos $p = 0,02$, por lo que $q = 1 - p = 0,98$, y por tanto $x = \frac{m-np}{\sqrt{npq}} = \frac{230-200}{14} = 2,14$. Por ésto

$$P_{10000}(230) \approx \frac{1}{14}\varphi(2,14) = \frac{1}{14} \times 0,0404 = 0,0029.$$

2.3. Teorema límite integral de De Moivre–Laplace

Consideremos el espacio de probabilidad $(\Omega_n, \mathcal{A}_n, \mathbf{P})$ correspondiente a una serie de n experimentos independientes, y sea $\mu(\omega)$ la cantidad de éxitos que ocurren en estos n experimentos. Queremos obtener una expresión aproximada, para la probabilidad de un suceso de la forma $\{\omega \in \Omega_n : \alpha < \mu(\omega) \leq \beta\}$, donde $\alpha < \beta$ son arbitrarios, como el que aparece en la parte (b) del ejercicio 2.3. Para ésto, consideremos la notación

$$\mathbf{P}(\alpha < \mu \leq \beta) = \mathbf{P}(\{\omega \in \Omega_n : \alpha < \mu(\omega) \leq \beta\}).$$

Teorema 2.2 (Teorema límite integral de De Moivre–Laplace).

Consideremos una serie de n experimentos independientes, con probabilidad de éxito en cada experimento igual a p ($0 < p < 1$), $q = 1 - p$. Sea μ la cantidad de éxitos que ocurren en esta serie. Sean a, b dos números que verifican $-\infty \leq a < b \leq \infty$ (es decir, incluimos las posibilidades $a = -\infty$ y $b = \infty$). Entonces, tiene lugar la convergencia

$$\mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \rightarrow 0, \quad \text{si } n \rightarrow \infty, \quad (2.8)$$

uniformemente, para todos los valores de a, b considerados.

Una manera alternativa de escribir la convergencia que tiene lugar en (2.8), es

$$\sup_{-\infty \leq a < b \leq \infty} \left| \mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \right| \rightarrow 0, \quad \text{si } n \rightarrow \infty.$$

La demostración de este teorema se basa en la aplicación del Teorema límite local de De Moivre–Laplace 2.1 y en el siguiente resultado.

Lema 2.1. *Es válida la desigualdad*

$$P_n(m) \leq P_n([(n+1)p]),$$

para todo m ($0 \leq m \leq n$). Aquí $[x]$ es la parte entera de x , es decir el mayor número entero que no supera a x .

Demostración del lema. Según la fórmula (2.2), tenemos

$$\begin{aligned} \frac{P_n(m+1)}{P_n(m)} &= \frac{\binom{n}{m+1} p^{m+1} q^{n-m-1}}{\binom{n}{m} p^m q^{n-m}} \\ &= \frac{p}{q} \times \frac{n! m! (n-m)!}{(m+1)! (n-m-1)! n!} = \frac{p}{q} \times \frac{n-m}{m+1}. \end{aligned}$$

La desigualdad $P_n(m+1) > P_n(m)$ es equivalente, entonces, a la desigualdad $(n-m)p > (m+1)q$, que se reescribe como $m < (n+1)p - 1$; la desigualdad $P_n(m+1) < P_n(m)$, equivalente a $m > (n+1)p - 1$. Si $(n+1)p$ no es un número natural, entonces $P_n(m)$ alcanza su valor máximo cuando $m = m_0 = [(n+1)p]$. Si $(n+1)p$ es un número natural, entonces el valor máximo de $P_n(m)$ se alcanza en dos valores de m : $m = m_0 = (n+1)p$, y $m = m_0 - 1$, teniéndose $P_n(m_0) = P_n(m_0 - 1)$. Como $m_0 = (n+1)p = [m_0(p+1)]$, la demostración está concluida. \square

Demostración del teorema 2.2. Introduzcamos las notaciones

$$P_n(a, b) = \mathbf{P} \left(a < \frac{\mu - np}{\sqrt{npq}} \leq b \right), \quad x_{n,m} = \frac{m - np}{\sqrt{npq}}.$$

Demostramos primero la convergencia en (2.8), cuando a y b son constantes finitas. Es clara la igualdad

$$P_n(a, b) = \sum_m P_n(m), \tag{2.9}$$

donde la suma se efectúa en los valores de m , para los cuales $a < x_{n,m} \leq b$. Tenemos $x_{n,m+1} - x_{n,m} = 1/\sqrt{npq}$, y además

$$\begin{aligned} x_{n,0} &= -\frac{np}{\sqrt{npq}} \rightarrow -\infty \quad (n \rightarrow \infty), \\ x_{n,n} &= \frac{n - np}{\sqrt{npq}} \rightarrow +\infty \quad (n \rightarrow \infty). \end{aligned}$$

Introducimos la función

$$\Pi_n(x) = \begin{cases} 0, & \text{si } x \leq x_{n,0}, \text{ ó } x > x_{n,n} + \frac{1}{\sqrt{npq}} \\ \sqrt{npq}P_n(m), & \text{si } x_{n,m} < x \leq x_{n,m+1} \quad (m = 0, 1, \dots, n). \end{cases}$$

y designamos mediante \underline{m} y \overline{m} , a los (únicos) naturales, que verifican las condiciones

$$a < x_{n,\underline{m}} \leq a + \frac{1}{\sqrt{npq}}, \quad b < x_{n,\overline{m}} \leq b + \frac{1}{\sqrt{npq}}.$$

Es claro, que

$$\int_{x_{n,m}}^{x_{n,m+1}} \Pi_n(x) dx = \sqrt{npq}P_n(m)(x_{n,m+1} - x_{n,m}) = P_n(m),$$

para $m = 0, 1, \dots, n$, y la igualdad (2.9) puede ser escrita, como

$$P_n(a, b) = \sum_m \int_{x_{n,m}}^{x_{n,m+1}} \Pi_n(x) dx = \int_{x_{n,\underline{m}}}^{x_{n,\overline{m}}} \Pi_n(x) dx. \quad (2.10)$$

Para concluir la demostración (con a y b constantes finitas, fijas), hay que aproximar a la suma a la derecha en (2.10), mediante la integral en el intervalo $[a, b]$ de la función $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Esto se hace en dos etapas: (a) se observa que la diferencia de integral de la función $\Pi_n(x)$ en los intervalos $[a, b]$ y $[x_{n,\underline{m}}, x_{n,\overline{m}}]$ es arbitrariamente pequeña, para valores grandes de n (esto se hace utilizando el lema 2.1); (b) se aproximan las integrales en el intervalo $[a, b]$ de las funciones $\Pi_n(x)$ y $\varphi(x)$ (aquí se utiliza el teorema 2.1).

Comenzamos entonces dividiendo en tres el intervalo de integración en (2.10), obteniendo

$$P_n(a, b) = \int_a^b \Pi_n(x) dx - \int_a^{x_{n,\underline{m}}} \Pi_n(x) dx + \int_b^{x_{n,\overline{m}}} \Pi_n(x) dx,$$

de donde resulta, que

$$\left| P_n(a, b) - \int_a^b \Pi_n(x) dx \right| \leq \int_a^{x_{n,\underline{m}}} \Pi_n(x) dx + \int_b^{x_{n,\overline{m}}} \Pi_n(x) dx, \quad (2.11)$$

Por el lema anterior, que proporciona una acotación para el integrando, se tiene $\Pi_n(x) \leq \sqrt{npq}P_n(m_0)$ para todo x , donde $m_0 = [(n+1)p]$. Como vale $(n+1)p - 1 < m_0 \leq (n+1)p$ tenemos, con $x_{n,m_0} = m_0 -$

np/\sqrt{npq} , por un lado, $x_{n,m_0} \leq (np + p - np)/\sqrt{npq}$, y por otro, $x_{n,m_0} \geq np + p - 1 - np/\sqrt{npq}$. Por ésto, para todo n suficientemente grande, se verifica $-1 \leq x_{n,m_0} \leq 1$, y por el teorema 2.1 (donde tomamos $C = 1$), tenemos

$$\frac{\sqrt{npq}P_n(m_0)}{\frac{1}{\sqrt{2\pi}}e^{-x_{n,m_0}^2/2}} \rightarrow 1 \quad \text{cuando } n \rightarrow \infty.$$

De aquí, como $\varphi(x) \leq 1/\sqrt{2\pi}$, obtenemos

$$\Pi_n(x) \leq \sqrt{npq}P_n(m_0) < \frac{2}{\sqrt{2\pi}}$$

para todo n suficientemente grande. Teniendo en cuenta esta acotación, y la fórmula (2.11), obtenemos

$$\begin{aligned} \left| P_n(a, b) - \int_a^b \Pi_n(x) dx \right| &\leq \sqrt{npq}P_n(m_0)(x_{n,\underline{m}} - a + x_{n,\overline{m}} - b) \\ &\leq \sqrt{npq}P_n(m_0) \frac{2}{\sqrt{npq}} < \frac{4}{\sqrt{2\pi npq}}, \end{aligned} \quad (2.12)$$

para todo n suficientemente grande (esto concluye la etapa (a)).

Demostremos ahora, que

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}(1 + r_n(x)), \quad (2.13)$$

donde el resto $r_n(x)$ converge uniformemente a cero, es decir

$$\sup_{a \leq x \leq b} |r_n(x)| \rightarrow 0 \quad (n \rightarrow \infty). \quad (2.14)$$

Sea x fijo en el intervalo $[a, b]$. Si $x_{n,m} < x \leq x_{n,m+1}$, por el teorema 2.1, tenemos

$$\Pi_n(x) = \sqrt{npq}P_n(m) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x_{n,m}^2}{2}}(1 + \gamma_n),$$

donde $\gamma_n \rightarrow 0$ ($n \rightarrow \infty$) uniformemente con respecto a m por el teorema 2.1. Sumando y restando en el exponente, tenemos

$$\Pi_n(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}e^{-(x^2 - x_{n,m}^2)/2}(1 + \gamma_n).$$

Es claro, que

$$\frac{|x^2 - x_{n,m}^2|}{2} = \frac{|(x - x_{n,m})(x + x_{n,m})|}{2} \leq \frac{1}{2\sqrt{npq}} 2 \max(|a|, |b|) \rightarrow 0,$$

si $n \rightarrow \infty$. Poniendo $r_n(x) = e^{-(x^2 - x_{n,m}^2)/2} (1 + \gamma_n) - 1$, la fórmula anterior es (2.13), y se verifica la convergencia uniforme en (2.14). Para concluir la etapa (b) utilizamos la acotación (2.12) y la fórmula (2.13), para obtener

$$\begin{aligned} & \left| P_n(a, b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \\ & \leq \left| P_n(a, b) - \int_a^b \Pi_n(x) dx \right| + \left| \int_a^b \Pi_n(x) dx - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \\ & \leq \frac{4}{\sqrt{2\pi npq}} + \frac{1}{\sqrt{2\pi}} \sup_{a \leq x \leq b} |r_n(x)| \int_a^b e^{-\frac{x^2}{2}} dx \rightarrow 0, \end{aligned}$$

si $n \rightarrow \infty$. La fórmula (2.8) está entonces demostrada, bajo el supuesto adicional, de que a y b son constantes finitas. Es fácil de ver en la demostración realizada, que esta condición puede sustituirse por la condición $-c \leq a < b \leq c$, donde c es una constante arbitraria, dado que se tiene

$$\sup_{a \leq x \leq b} |r_n(x)| \leq \sup_{-c \leq x \leq c} |r_n(x)| \rightarrow 0 \quad (n \rightarrow \infty).$$

Demostremos ahora el teorema 2.2 sin supuestos adicionales sobre a y b . Tiene lugar la igualdad

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}. \quad (2.15)$$

Sea ε un número positivo arbitrario. Sea c una constante positiva que verifica la condición

$$\frac{1}{\sqrt{2\pi}} \int_{\{|x|>c\}} e^{-\frac{x^2}{2}} dx < \varepsilon. \quad (2.16)$$

Según se vió en la primera parte de la demostración, tiene lugar la acotación

$$\left| P_n(-c, c) - \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx \right| < \varepsilon. \quad (2.17)$$

si n es suficientemente grande. Además,

$$\begin{aligned} \mathbf{P}\left(\frac{\mu - np}{\sqrt{npq}} \leq -c\right) + \mathbf{P}\left(\frac{\mu - np}{\sqrt{npq}} > c\right) &= 1 - P_n(-c, c) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-c}^c e^{-\frac{x^2}{2}} dx - P_n(-c, c) + \frac{1}{\sqrt{2\pi}} \int_{\{|x|>c\}} e^{-\frac{x^2}{2}} dx < 2\varepsilon. \end{aligned} \quad (2.18)$$

para todo n suficientemente grande, en vista de la elección de c (fórmula 2.16) y la acotación (2.17). Consideremos el caso en el que $a \leq -c < b \leq c$ (los otros dos casos a considerar son análogos). Tenemos

$$\begin{aligned} &\left| P_n(a, b) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \right| \\ &= \left| P_n(a, -c) + P_n(-c, b) - \frac{1}{\sqrt{2\pi}} \int_a^{-c} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi}} \int_{-c}^b e^{-\frac{x^2}{2}} dx \right| \\ &\leq \left| P_n(a, -c) \right| + \left| \frac{1}{\sqrt{2\pi}} \int_a^{-c} e^{-\frac{x^2}{2}} dx \right| + \left| P_n(-c, b) - \frac{1}{\sqrt{2\pi}} \int_{-c}^b e^{-\frac{x^2}{2}} dx \right| \\ &\qquad\qquad\qquad < 2\varepsilon + \varepsilon + \varepsilon = 4\varepsilon, \end{aligned}$$

para todo n suficientemente grande, teniendo en cuenta las acotaciones (2.18), (2.16) y (2.17). Esto concluye la demostración del teorema. \square

Del teorema 2.2 se obtiene, que para n suficientemente grande, tiene lugar la identidad aproximada

$$\mathbf{P}\left(a < \frac{\mu - np}{\sqrt{npq}} \leq b\right) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a),$$

si introducimos la función

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad (2.19)$$

definida para todo x real. Esta función se denomina *función de distribución normal*. En el final del libro se presenta una tabla con los valores de la función $\Phi(x)$.

En diversas aplicaciones surge, en forma frecuente, la necesidad de calcular probabilidades de la forma $\mathbf{P}(\alpha \leq \mu \leq \beta)$, para α y β dados. Es claro, que

$$\mathbf{P}(\alpha \leq \mu \leq \beta) = \mathbf{P}\left(\frac{\alpha - np}{\sqrt{npq}} \leq \frac{\mu - np}{\sqrt{npq}} \leq \frac{\beta - np}{\sqrt{npq}}\right),$$

de donde

$$\mathbf{P}(\alpha \leq \mu \leq \beta) \approx \Phi\left(\frac{\beta - np}{\sqrt{npq}}\right) - \Phi\left(\frac{\alpha - np}{\sqrt{npq}}\right).$$

Una probabilidad de este tipo, es, por ejemplo, la que aparece en la parte (b) del ejemplo 2.3, en el cual se requería calcular $\mathbf{P}(\mu \leq 230)$. En este caso, se tiene $p = 0,02$, $q = 1 - p = 0,98$, $\alpha = 0$, $\beta = 230$, $n = 10000$, de forma que

$$\mathbf{P}(\mu \leq 230) \approx \Phi\left(\frac{230 - 200}{14}\right) - \Phi\left(\frac{-200}{14}\right) \approx \Phi(2,14) = 0,9838.$$

Observemos que, habitualmente, las tablas de la función $\Phi(x)$ incluyen únicamente valores positivos de x . Para determinar $\Phi(x)$ con $x < 0$, se utiliza la igualdad $\Phi(-x) = 1 - \Phi(x)$, válida para todo x . Esta fórmula se deduce de (2.19). En el capítulo 3 se considerarán funciones de distribución normales de un tipo más general, dependientes de dos parámetros.

2.4. Teorema de Bernoulli

Como corolario del teorema integral de De Moivre–Laplace, se obtiene el siguiente resultado.

Teorema 2.3 (Teorema de Bernoulli).

Consideremos una serie de n experimentos independientes, con probabilidad de éxito en cada experimento igual a p ($0 < p < 1$), $q = 1 - p$. Sea μ la cantidad de éxitos que ocurren en esta serie. Entonces, para todo $\varepsilon > 0$ fijo, tiene lugar la convergencia

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \rightarrow 1 \quad \text{si } n \rightarrow \infty. \quad (2.20)$$

Demostración. Tenemos

$$\begin{aligned} \mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) &= \mathbf{P}\left(\left\{\omega \in \Omega_n : \left|\frac{\mu(\omega)}{n} - p\right| < \varepsilon\right\}\right) \\ &= \mathbf{P}\left(-\varepsilon < \frac{\mu - np}{n} < \varepsilon\right) \\ &= \mathbf{P}\left(-\varepsilon\sqrt{\frac{n}{pq}} < \frac{\mu - np}{\sqrt{npq}} < \varepsilon\sqrt{\frac{n}{pq}}\right), \end{aligned}$$

por lo que, en vista de la identidad (2.15), obtenemos $\mathbf{P}(|\mu/n - p| < \varepsilon) - 1 = R_1 - R_2$, donde

$$R_1 = \mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) - \frac{1}{\sqrt{2\pi}} \int_{\{|x| < \varepsilon\sqrt{\frac{n}{pq}}\}} e^{-x^2/2} dx,$$

$$R_2 = \frac{1}{\sqrt{2\pi}} \int_{\{|x| > \varepsilon\sqrt{\frac{n}{pq}}\}} e^{-x^2/2} dx.$$

Sea δ arbitrario y positivo. Por el Teorema integral de De Moivre–Laplace, se obtiene, que $|R_1| < \delta$ si $n \geq n_1$ para algún n_1 (aquí se utiliza la convergencia uniforme). Es claro también, que $|R_2| < \delta$ si $n \geq n_2$, para algún n_2 . Entonces,

$$\left|\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) - 1\right| \leq |R_1| + |R_2| < 2\delta$$

para $n \geq \max(n_1, n_2)$. De aquí se obtiene la convergencia en (2.20). \square

De la demostración del teorema 2.3 resulta también, que para n suficientemente grande, vale la identidad aproximada

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \approx \int_{\{|x| < \varepsilon\sqrt{\frac{n}{pq}}\}} e^{-x^2/2} dx = \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right),$$

la cual, utilizando la identidad $\Phi(-x) = 1 - \Phi(x)$, se escribe como

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - 1.$$

La proporción μ/n es la frecuencia de éxitos en n experimentos. Como consecuencia del teorema de Bernoulli tenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - p\right| < \varepsilon\right) \approx 1$$

para cualquier $\varepsilon > 0$ arbitrariamente pequeño, si la cantidad de experimentos es suficientemente grande.

Un suceso cuya probabilidad es cercana a la unidad se dice *prácticamente seguro*. El resultado obtenido puede ser entonces enunciado de la forma siguiente: *si la cantidad de experimentos n es suficientemente grande, es prácticamente seguro que la diferencia entre la frecuencia de*

éxitos en n experimentos y la probabilidad de éxito de un experimento sea arbitrariamente pequeña.

El teorema de Bernoulli es la forma más sencilla de un conjunto de resultados relativos a la convergencia de frecuencias o de promedios, que se denominan *leyes de los grandes números*. En el capítulo 5 estudiaremos teoremas más generales de este tipo.

2.5. Aproximación de Poisson a la distribución binomial

El siguiente resultado es una aproximación diferente de las estudiadas, para la probabilidad de que ocurran una cantidad determinada de éxitos en una serie de experimentos independientes, especialmente útil cuando la probabilidad de éxito es pequeña (y la cantidad de experimentos grande).

Proposición 2.2. *Consideremos, para cada n natural, una serie de n experimentos independientes, con probabilidad de éxito en cada experimento igual a λ/n ; la constante λ es positiva y arbitraria. Sea μ_n la cantidad de éxitos que ocurren en la serie n -ésima. Entonces, tiene lugar la convergencia*

$$\mathbf{P}(\mu_n = m) \rightarrow \frac{\lambda^m}{m!} e^{-\lambda} \quad \text{si } n \rightarrow \infty.$$

Demostración. Según (2.2), tenemos

$$\begin{aligned} \mathbf{P}(\mu_n = m) &= \binom{n}{m} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\ &= \frac{\lambda^m}{m!} \times \frac{n(n-1)\cdots(n-m+1)}{n^m} \left(1 - \frac{\lambda}{n}\right)^{n-m} \\ &= \frac{\lambda^m}{m!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \\ &\quad \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-m} \rightarrow \frac{\lambda^m}{m!} e^{-\lambda}, \end{aligned}$$

si $n \rightarrow \infty$, dado que el primer factor es constante (no depende de n), para el segundo factor tenemos $(1 - \lambda/n)^n \rightarrow e^{-\lambda}$ ($n \rightarrow \infty$), y los restantes m factores convergen a 1. \square

Según la proposición recién demostrada, si n es suficientemente grande, tiene lugar la identidad aproximada

$$\mathbf{P}(\mu_n = m) \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

que, alternativamente, se escribe como

$$\mathbf{P}(\mu_n = m) \approx \frac{(np)^m}{m!} e^{-np}, \quad (2.21)$$

donde p designa la probabilidad de éxito en un experimento. La fórmula de aproximación (2.21) es adecuada, en aquellos casos en que la cantidad de experimentos n es grande y la probabilidad p de éxito de cada experimento es pequeña.

Ejemplo 2.4. La probabilidad de acertar en un blanco en cada disparo es de 0,01. Calcular la probabilidad de que ocurra, por lo menos, un acierto en 400 disparos.

Solución. Tenemos $\mathbf{P}(\mu_{400} = 0) \approx e^{-400(0,01)} = e^{-4} = 0,0183$, por lo que

$$\mathbf{P}(\mu_{400} \geq 1) = 1 - \mathbf{P}(\mu_{400} = 0) \approx 0,9817.$$

2.6. Ejercicios

1. Se tira una moneda 6 veces consecutivas. Calcular la probabilidad de que aparezca cara: (a) por lo menos una vez; (b) no menos de dos veces; (c) de 3 a 5 veces.
2. Calcular la probabilidad de obtener tres veces 6 puntos, al tirar un dado 5 veces.
3. En un proceso industrial, la probabilidad de que un cierto artículo resulte defectuoso es 0,01. Calcular la probabilidad de que, en 10 artículos elegidos al azar, resulten: (a) por lo menos un defectuoso; (b) no menos de dos defectuosos.
4. En la transmisión de un mensaje compuesto por signos, la probabilidad de que ocurra un error en un signo es 0,1. Calcular la probabilidad de que, en un mensaje con 4 signos: (a) no hayan errores; (b) ocurra un error; (c) ocurra no menos de un error.

5. Calcular la probabilidad de que, en $2n$ experimentos en un esquema de Bernoulli, se obtengan éxitos únicamente en los n experimentos con número par, si la probabilidad de éxito en un experimento es p .
6. Un trabajador controla 5 máquinas de un mismo tipo. La probabilidad de que una máquina requiera la atención del trabajador en el lapso de una hora es $1/3$. Calcular la probabilidad de que, en el curso de una hora, el trabajador sea requerido por: (a) 2 máquinas; (b) no menos de 2 máquinas.
7. Un matemático lleva consigo dos cajas de fósforos. Al principio en cada caja hay n fósforos. Cada vez que el matemático precisa un fósforo, elige al azar una de las cajas. Calcular la probabilidad de que, cuando el matemático encuentre una caja vacía, en la otra hayan exactamente r fósforos ($0 < r \leq n$).
8. En una habitación hay tres lámparas. La probabilidad de que cada una de estas lámparas no se quemé, en el lapso de un año, es 0,8. Calcular la probabilidad de que, en el curso de un año, estén funcionando: (a) 2 lámparas; (b) por lo menos una lámpara.
9. La probabilidad de éxito en un esquema de Bernoulli es p . Calcular la probabilidad de que, en el experimento que ocupa el k -ésimo lugar, ocurra éxito por ℓ -ésima vez ($0 < \ell \leq k \leq n$).
10. Una partícula que fluctúa por los puntos enteros de la recta real, en un cierto momento (momento de salto) se traslada una unidad a la izquierda con probabilidad $1/2$, o una unidad a la derecha con probabilidad $1/2$ (independientemente de la dirección de los movimientos anteriores). Este esquema se denomina *paseo al azar simple*. Calcular la probabilidad de que, luego de $2n$ saltos, la partícula se encuentre en el punto desde el cual comenzó a trasladarse.
11. Se tira una moneda 1600 veces. Calcular aproximadamente, la probabilidad de que se obtenga cara: (a) exactamente 780 veces; (b) de 780 a 820 veces.
12. La probabilidad de acertar en un blanco es 0,8. Calcular aproximadamente, la probabilidad de que en 400 disparos, se obtengan: (a) exactamente 300 aciertos; (b) no menos de 300 aciertos.

- 13.** En determinadas condiciones de producción de un cierto artículo, la probabilidad de que resulte defectuoso es 0,01. Calcular la probabilidad de que, entre 10000 artículos examinados de esta producción, resulten: (a) de 80 a 110 defectuosos; (b) no menos de 9950 artículos sin defectos.
- 14.** En una compañía de seguros hay asegurados 50.000 personas, de una cierta edad y grupo social. La probabilidad de defunción en el curso de un año, para cada individuo, es 0,006. Cada persona asegurada paga, al inicio del año, 40 dólares, y en caso de fallecer, sus parientes reciben de la compañía 5000 dólares. Calcular la probabilidad de que, en el lapso de un año, dicha compañía: (a) sufra pérdidas; (b) obtenga ganancias de por lo menos 300.000 dólares; (c) obtenga ganancias de por lo menos 800.000 dólares.
- 15.** Calcular la probabilidad de que, en una serie de 1000 tiradas de una moneda, la frecuencia de aparición de cara se diferencie de la probabilidad de aparición de cara, en no más de 0,03.
- 16.** La probabilidad de éxito en un esquema de Bernoulli es 0,005. Calcular la probabilidad de que, en una serie de 800 experimentos, ocurra por lo menos un éxito. (Sugerencia: utilizar la aproximación de Poisson a la distribución binomial.)
- 17.** La probabilidad de acertar en un blanco es de 0,001. Calcular la probabilidad de acertar en el blanco dos o más veces, en una serie de 5000 disparos. (Sugerencia: utilizar la aproximación de Poisson a la distribución binomial.)

Capítulo 3

VARIABLES ALEATORIAS Y DISTRIBUCIONES DE PROBABILIDAD

3.1. Variables aleatorias y funciones de distribución

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. Llamamos *variable aleatoria* a una función $X = X(\omega)$ que toma valores reales, definida en el espacio de sucesos elementales Ω , y que verifica la condición

$$\{\omega \in \Omega: X(\omega) \leq x\} \in \mathcal{A} \tag{3.1}$$

para todo x real.

En la terminología del análisis real, una función $X(\omega)$ que cumple la condición (3.1) para todo x , se denomina *medible*. De esta forma, una variable aleatoria es una función real y medible de los sucesos elementales. Se puede verificar que la condición (3.1) para todo x , es equivalente a la condición

$$\{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{A} \tag{3.2}$$

para cualquier conjunto boreliano¹ B de puntos de la recta real \mathbb{R} . En el caso particular en el que B es el intervalo $(-\infty, x]$, la condición (3.2) se convierte en la condición (3.1).

Veamos ahora algunos ejemplos de variables aleatorias.

¹La clase de los conjuntos borelianos en la recta es la mínima σ -álgebra de conjuntos, que contiene a todos los intervalos.

Ejemplo 3.1. Consideremos un espacio de sucesos elementales Ω compuesto por una cantidad finita o numerable de puntos $\omega_1, \omega_2, \dots$. Sea \mathcal{A} el conjunto de todos los subconjuntos de Ω . (Es claro que este conjunto es una σ -álgebra.) Introducimos las probabilidades de los sucesos, asignándole a cada suceso elemental ω_i , en calidad de probabilidad, un número no negativo p_i ($i = 1, 2, \dots$), de forma tal que la suma total de las probabilidades asignadas sea 1, es decir, $p_1 + p_2 + \dots = 1$.

En este espacio, cualquier función real $X = X(\omega)$, $\omega \in \Omega$, resulta ser una variable aleatoria, porque la condición (3.1) se verifica automáticamente para cualquier x , en vista de nuestra elección de la σ -álgebra \mathcal{A} . Entonces, en la situación considerada, la definición de variable aleatoria es muy sencilla: una variable aleatoria es una función que a cada suceso elemental le hace corresponder un número real.

Ejemplo 3.2. Consideremos un espacio de sucesos elementales Ω compuesto por seis puntos $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ y ω_6 . Sea \mathcal{A} la σ -álgebra formada por todos los subconjuntos de Ω . Asignamos a cada uno de estos 6 puntos del espacio Ω la misma probabilidad, es decir, $1/6$. Como observamos en la sección 1.2, el espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ así construido es un modelo matemático del experimento consistente en tirar un dado equilibrado. Convenimos en que el punto ω_i en este modelo está numerado de forma tal, que corresponde a la aparición de i puntos en la cara superior del dado. La función $X(\omega_i) = i$ ($i = 1, \dots, 6$), que podemos interpretar como la cantidad de puntos obtenida luego de tirar el dado, es una variable aleatoria. (En el ejemplo anterior se consideró una clase mas general de variables aleatorias.)

Ejemplo 3.3. Consideremos el espacio de sucesos elementales Ω compuesto por los puntos del intervalo $[0, 1]$; sean \mathcal{A} la σ -álgebra de los conjuntos borelianos de este intervalo, y \mathbf{P} la medida de Lebesgue en este intervalo (ver ejemplo 1.3). La terna $(\Omega, \mathcal{A}, \mathbf{P})$ es un espacio de probabilidad, y cualquier función boreliana $X = X(\omega)$, $\omega \in [0, 1]$, es una variable aleatoria².

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega)$, $\omega \in \Omega$. Como el conjunto $\{\omega \in \Omega : X(\omega) \leq x\}$ es un suceso (es decir, un conjunto de la σ -álgebra de sucesos \mathcal{A}), está definida la probabilidad $\mathbf{P}(\{\omega : X(\omega) \leq x\})$ para todo $x \in \mathbb{R}$; esta probabilidad será designada por brevedad $\mathbf{P}(X \leq x)$ (se lee: la probabilidad

²Una función real X definida en un intervalo \mathbf{I} se llama *boreliana*, si $\{\omega \in \mathbf{I} : X(\omega) \leq x\}$ es un conjunto de la σ -álgebra de Borel en \mathbf{I} para todo x real.

de que la variable aleatoria X tome un valor menor o igual que x). Se denomina *función de distribución* de la variable aleatoria X , a la función $F(x)$, definida para todos los valores x reales, mediante la fórmula

$$F(x) = \mathbf{P}(X \leq x). \quad (3.3)$$

Observemos que si X es una variable aleatoria y B un conjunto boreliano de puntos de la recta real \mathbb{R} , está definida la probabilidad $\mathbf{P}(\{\omega: X(\omega) \in B\})$, que será designada por $\mathbf{P}(X \in B)$, también por brevedad. La función $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$, definida para todos los conjuntos borelianos B de puntos de la recta real, se llama *función de probabilidad* de la variable aleatoria X . Es claro que $\mathbf{P}_X((-\infty, x]) = \mathbf{P}(X \leq x) = F(x)$ para cualquier x , donde $F(x)$ es la función de distribución de la variable aleatoria X . Llamamos *distribución de probabilidad* (o más sencillamente *distribución*) de la variable aleatoria X , indistintamente, a la función de distribución $F(x)$ de la variable aleatoria X , o a la función de probabilidad $\mathbf{P}_X(B)$ de esta variable aleatoria.

Lema 3.1. *Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega)$, $\omega \in \Omega$. Sean dos números $a < b$. Entonces, los conjuntos $\{\omega: X(\omega) < a\}$, $\{\omega: X(\omega) = a\}$, $\{\omega: a < X(\omega) < b\}$, $\{\omega: a < X(\omega) \leq b\}$, $\{\omega: a \leq X(\omega) < b\}$ y $\{\omega: a \leq X(\omega) \leq b\}$ son sucesos (es decir, elementos de la σ -álgebra \mathcal{A}).*

Demostración. Estas afirmaciones son inmediatas, si consideramos conocida la equivalencia entre las condiciones (3.1) y (3.2). En caso contrario, la demostración puede basarse en otros hechos conocidos del análisis real, pudiéndose también dar una demostración directa, cosa que haremos a continuación.

Como X es una variable aleatoria, el conjunto $\{\omega: X(\omega) \leq a - 1/n\}$ pertenece a \mathcal{A} para cualquier a real y cualquier n natural. Por ésto, en vista de la definición de σ -álgebra, la suma $\bigcup_{n=1}^{\infty} \{\omega: X(\omega) \leq a - 1/n\}$ también pertenece a \mathcal{A} . La suma de sucesos anterior es igual al conjunto $\{\omega: X(\omega) < a\}$ que, por lo tanto, es un suceso y pertenece a la σ -álgebra \mathcal{A} . Los restantes enunciados se demuestran en forma análoga. \square

En vista del lema recién demostrado están definidas las probabilidades $\mathbf{P}(X < a)$, $\mathbf{P}(a \leq X < b)$, $\mathbf{P}(X = a)$, etc. para cualquier variable aleatoria X y reales $a < b$ arbitrarios.

Estudiemos ahora las propiedades que verifica la función de distribución $F(x)$ de una variable aleatoria X .

Propiedad 1. Se verifica $0 \leq F(x) \leq 1$, para todo x real.

Esta afirmación es consecuencia inmediata de la definición (3.3).

Propiedad 2. Si $a < b$, entonces $\mathbf{P}(a < X \leq b) = F(b) - F(a)$.

Demostración. Observemos, que

$$\{\omega: X(\omega) \leq b\} = \{\omega: X(\omega) \leq a\} \cup \{\omega: a < X(\omega) \leq b\},$$

donde los sucesos a la derecha son incompatibles. Entonces, aplicando el axioma III, obtenemos

$$\mathbf{P}(X \leq b) = \mathbf{P}(X \leq a) + \mathbf{P}(a < X \leq b).$$

De aquí, y de la fórmula (3.3), se concluye la validez de la propiedad. \square

Propiedad 3. La función $F(x)$ es no decreciente en toda la recta real, es decir, dados $a < b$ reales, vale $F(a) \leq F(b)$.

Demostración. Esta afirmación se deduce de la propiedad anterior, teniendo en cuenta que, si $a < b$, entonces

$$F(b) - F(a) = \mathbf{P}(a < X \leq b) \geq 0.$$

es decir, $F(a) \leq F(b)$. \square

Propiedad 4. Se tiene $\lim_{x \rightarrow +\infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$.

Demostración. Es claro que $\Omega = \bigcup_{m=-\infty}^{m=+\infty} \{\omega: m-1 < X(\omega) \leq m\}$. Entonces, como los sucesos que aparecen en la suma son incompatibles dos a dos, aplicando el axioma III, tenemos

$$\begin{aligned} 1 &= \mathbf{P}(\Omega) = \sum_{m=-\infty}^{m=+\infty} \mathbf{P}(\{\omega: m-1 < X(\omega) \leq m\}) \\ &= \lim_{N \rightarrow \infty} \sum_{m=-N}^{m=N} \mathbf{P}(m-1 < X(\omega) \leq m) \\ &= \lim_{N \rightarrow \infty} \sum_{m=-N}^{m=N} (F(m) - F(m-1)) = \lim_{N \rightarrow \infty} (F(N) - F(-N-1)). \end{aligned}$$

Como existen los límites $\lim_{x \rightarrow \infty} F(x)$, $\lim_{x \rightarrow -\infty} F(x)$, porque las funciones de distribución son no decrecientes (propiedad 3), de la igualdad obtenida y de la propiedad 1, obtenemos (como única posibilidad), que $\lim_{x \rightarrow \infty} F(x) = 1$, $\lim_{x \rightarrow -\infty} F(x) = 0$, concluyendo la demostración. \square

Propiedad 5. *La función de distribución $F(x)$ es continua por la derecha.*

Demostración. Sea x un real arbitrario. Consideremos una sucesión numérica x_1, x_2, \dots decreciente y convergente a x . Es decir $x_1 > x_2 > \dots$, $x_n \rightarrow x$ ($n \rightarrow \infty$). Sea $\mathbf{A}_n = \{\omega: x < X \leq x_n\}$. Es claro que $\mathbf{A}_1 \supset \mathbf{A}_2 \supset \dots$, y que $\bigcap_{n=1}^{\infty} \mathbf{A}_n = \emptyset$. Por la propiedad 8 en la sección 1.3, existe el $\lim_n \mathbf{P}(\mathbf{A}_n) = \mathbf{P}(\emptyset) = 0$. Entonces, aplicando la propiedad 3, obtenemos $\mathbf{P}(\mathbf{A}_n) = \mathbf{P}(x < X \leq x_n) = F(x_n) - F(x) \rightarrow 0$ ($n \rightarrow \infty$), lo que concluye la demostración. \square

Propiedad 6. *Una función de distribución tiene una cantidad finita o numerable de puntos de discontinuidad.*

Demostración. Por las propiedades 1 y 3, la función $F(x)$ tiene:

- a lo sumo un salto de magnitud h , con $h > 1/2$,
- a lo sumo dos saltos de magnitud h , con $1/2 \geq h > 1/3$,
- a lo sumo tres saltos de magnitud h , con $1/3 \geq h > 1/4$,
- \vdots
- a lo sumo m saltos de magnitud h , con $1/m \geq h > 1/(m+1)$,
- \vdots

En consecuencia, el conjunto de los puntos de discontinuidad de la función $F(x)$ es finito o numerable, dado que sus elementos se pueden numerar. \square

3.2. Variables aleatorias con distribuciones discretas y absolutamente continuas

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega)$, $\omega \in \Omega$. Decimos que la variable aleatoria X tiene *distribución discreta* si existe un conjunto B finito o numerable de puntos de la recta real, tal que se verifica $\mathbf{P}(X \in B) = 1$.

Si X es una variable aleatoria discreta, y un punto x verifica $p = \mathbf{P}(X = x) > 0$, decimos que la variable aleatoria X *toma el valor x* con probabilidad p .

Consideremos entonces una variable aleatoria X , que toma los valores x_1, x_2, \dots , con probabilidades p_1, p_2, \dots , respectivamente; es decir, $p_k = \mathbf{P}(X = x_k)$ ($k = 1, 2, \dots$) con $p_1 + p_2 + \dots = 1$. Utilizando esta información, no es difícil calcular los valores que toma la función de distribución $F(x)$ de esta variable aleatoria X , en cada valor real x . En efecto, los sucesos $\{\omega: X(\omega) \leq x\}$ y $\bigcup_{k: x_k \leq x} \{\omega: X(\omega) = x_k\}$ tienen la misma probabilidad (donde la suma de sucesos se efectúa para todos los k tales que $x_k \leq x$). Entonces $\mathbf{P}(X \leq x) = \sum_{k: x_k \leq x} \mathbf{P}(X = x_k)$, lo que significa, que

$$F(x) = \sum_{k: x_k \leq x} p_k. \quad (3.4)$$

De (3.4) se deduce, que el gráfico de la función $F(x)$ es constante, en los intervalos delimitados por dos valores consecutivos que toma la variable aleatoria, siendo estos valores puntos de salto. La magnitud del salto en el cada punto x_k es igual a p_k .

Veamos algunos ejemplos importantes de distribuciones discretas.

Ejemplo 3.4. Decimos que una variable aleatoria tiene *distribución degenerada* si existe un número c tal que $\mathbf{P}(X = c) = 1$. La función de distribución $F(x)$ de esta variable aleatoria vale 0 si $x < c$, y vale 1 si $x \geq c$.

Ejemplo 3.5. Decimos que una variable aleatoria tiene *distribución binomial* con parámetros (n, p) , donde n es un natural y $0 < p < 1$, si se verifica

$$\mathbf{P}(X = m) = \binom{n}{m} p^m (1 - p)^{n-m} \text{ para } m = 0, 1, \dots, n.$$

Veamos, como ejemplo, el gráfico de la función de distribución $F(x)$ de una variable aleatoria con distribución binomial, con parámetros $(2, 1/3)$ (ver figura 3.1). En el caso particular en el que $n = 1$, obtenemos la *distribución de Bernoulli*. De esta forma, una variable aleatoria tiene distribución de Bernoulli cuando toma dos valores: el valor 0 con probabilidad p , y el valor 1 con probabilidad $1 - p$. Como se estudió en el capítulo 2, la distribución binomial con parámetros (n, p) corresponde a una variable aleatoria que cuenta el número de éxitos en una serie n experimentos independientes, con probabilidad de éxito en cada experimento igual a p .

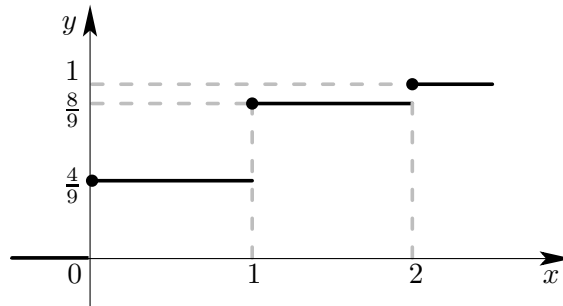


Figura 3.1: Gráfico de la función $y = F(x)$ para una variable aleatoria con distribución binomial, con parámetros $(2, 1/3)$

Ejemplo 3.6. Decimos que una variable aleatoria X tiene *distribución de Poisson* con parámetro $\lambda > 0$, si se verifica

$$\mathbf{P}(X = m) = \frac{\lambda^m}{m!} e^{-\lambda} \text{ para } m = 0, 1, 2, \dots$$

Es claro que la asignación de probabilidades es correcta, porque

$$\sum_{m=0}^{\infty} \mathbf{P}(X = m) = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} e^{\lambda} = 1.$$

Hasta el momento, hemos considerado únicamente variables aleatorias que toman una cantidad finita o numerable de valores. Veamos ahora otro tipo de variables aleatorias.

Decimos que una variable aleatoria X tiene distribución *absolutamente continua*, cuando su función de distribución $F(x)$ puede representarse de la forma

$$F(x) = \int_{-\infty}^x p(u) du \quad (3.5)$$

para todo x real, donde $p(u)$ es una función no negativa e integrable³⁴.

³Observemos, que en análisis real, una función $F(x)$ que se representa de la forma (3.5) se denomina *absolutamente continua*. Cualquier función absolutamente continua es continua en todos los puntos. La afirmación recíproca, en general, es falsa. La integral a la derecha en (3.5) es la integral de Lebesgue.

⁴Para el lector no familiarizado con la teoría de la medida, asumiremos, que la función $p(u)$ es continua en todos los puntos, con excepción de una cantidad finita.

La función $p(u)$ se denomina *densidad* de la distribución de la variable aleatoria X , y decimos, por brevedad, que la variable aleatoria X tiene densidad dada por $p(x)$. Es claro que

$$\int_{-\infty}^{\infty} p(u)du = 1 \quad (3.6)$$

por (3.5) y la propiedad $\lim_{x \rightarrow \infty} F(x) = 1$. Si $p(u)$ es continua en un punto x , en este punto existe la derivada $F'(x) = p(x)$. En particular, si $p(u)$ es continua en todos los puntos, tenemos $F'(x) = p(x)$ para todo x .

No es difícil demostrar que si una variable aleatoria X tiene función de distribución absolutamente continua, entonces $\mathbf{P}(X = x_0) = 0$ para cualquier x_0 . En efecto, la función de distribución $F(x)$ es continua en todos los puntos en vista de (3.5), y por ésto, para cualquier $h > 0$, tenemos

$$0 \leq \mathbf{P}(X = x_0) \leq \mathbf{P}(x_0 - h < X \leq x_0) = F(x_0) - F(x_0 - h) \rightarrow 0,$$

si $h \rightarrow 0$, obteniendo que $\mathbf{P}(X = x_0) = 0$. Observemos una consecuencia inmediata de la proposición recién demostrada. Si una variable aleatoria X tiene distribución absolutamente continua con densidad $p(x)$, para dos números $a < b$ arbitrarios, tenemos

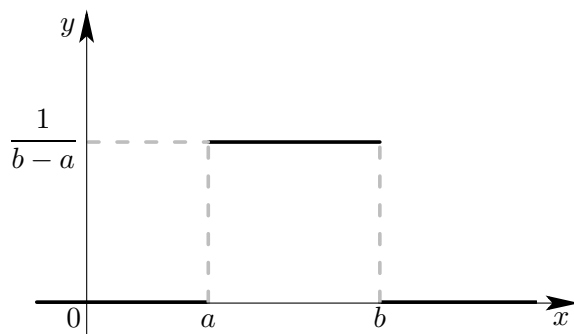
$$\begin{aligned} \mathbf{P}(a \leq X < b) &= \mathbf{P}(a \leq X \leq b) = \mathbf{P}(a < X \leq b) \\ &= \mathbf{P}(a < X < b) = \int_a^b p(x)dx. \end{aligned} \quad (3.7)$$

Veamos algunos ejemplos importantes de distribuciones absolutamente continuas.

Ejemplo 3.7. Decimos que una variable aleatoria X tiene *distribución uniforme* en el intervalo (a, b) , donde $a < b$, si tiene densidad dada por

$$p(x) = \begin{cases} c, & \text{si } a < x < b, \\ 0, & \text{si } x \leq a \text{ ó } x \geq b \end{cases} \quad (3.8)$$

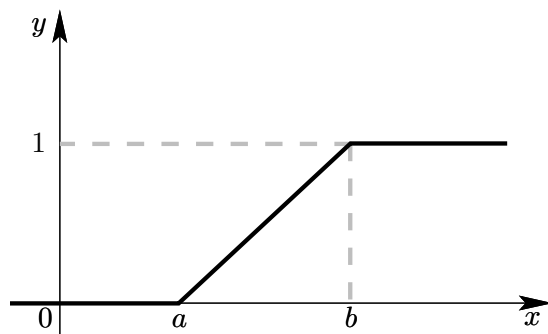
donde c es una constante. Esta constante se puede hallar utilizando la fórmula (3.6), válida para cualquier función de densidad. Para la función $p(x)$ dada por la fórmula (3.8), obtenemos $\int_a^b cdx = 1$, y de allí $c = 1/(b - a)$ (ver figura 3.2). Dada la densidad en (3.5), es fácil de obtener la

Figura 3.2: Gráfico de la densidad uniforme $p(x)$

función de distribución $F(x)$ de la variable aleatoria considerada. Tenemos

$$F(x) = \begin{cases} 0, & \text{si } x \leq a, \\ (x - a)/(b - a), & \text{si } a < x < b, \\ 1, & \text{si } x \geq b. \end{cases}$$

Esta función es continua para todo x , y es diferenciable en todos los puntos, con excepción de a y b (ver figura 3.3).

Figura 3.3: Gráfico de la función de distribución uniforme $F(x)$

Ejemplo 3.8. Decimos que una variable aleatoria X tiene *distribución exponencial* con parámetro $\alpha > 0$, si tiene densidad dada por

$$p(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases} \quad (3.9)$$

Es fácil de ver, que se verifica la igualdad (3.6), y que se cumple $p(x) \rightarrow 0$ ($x \rightarrow +\infty$), de forma que el eje Ox es asíntota del gráfico de $p(x)$. Dicho gráfico se representa en la figura 3.4.

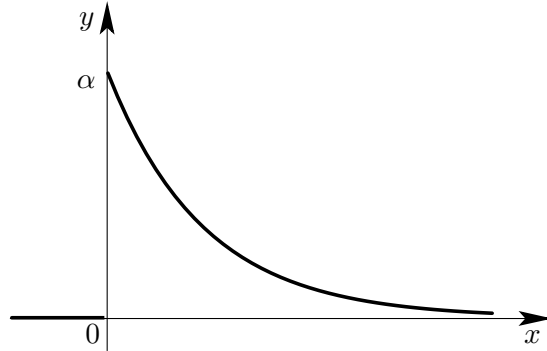


Figura 3.4: Gráfico de la densidad exponencial con parámetro α

Ejemplo 3.9. Decimos que una variable aleatoria tiene *distribución de Cauchy*, si tiene densidad dada por

$$p(x) = \frac{1}{\pi(1+x^2)}, \text{ para } x \text{ real.}$$

Ejemplo 3.10. Decimos que una variable aleatoria X tiene *distribución t* con n grados de libertad, donde $n \geq 1$ es un natural, si tiene densidad dada por

$$p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \text{ para } x \text{ real,} \quad (3.10)$$

donde la función

$$\Gamma(\lambda) = \int_0^{\infty} x^{\lambda-1} e^{-x} dx, \quad \lambda > 0. \quad (3.11)$$

se denomina *función Gama*. La densidad dada en (3.10) es simétrica respecto del eje Oy , y tiene un único máximo en el punto $x = 0$, mientras que el eje Ox resulta ser asíntota, si $|x| \rightarrow \infty$. Si $n = 1$ obtenemos la distribución de Cauchy.

La distribución de probabilidades absolutamente continua más importante es la *distribución normal*.

Ejemplo 3.11. Decimos que una variable aleatoria X tiene *distribución normal* con parámetros (a, σ) , donde a y $\sigma > 0$ son números reales, si tiene densidad dada por

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)} \text{ para } x \text{ real.} \quad (3.12)$$

Para verificar la fórmula (3.6) hay que hacer el cambio de variable $u = (x - a)/\sigma$ en la integral, y utilizar la identidad $\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$.

Es sencillo de ver que la función $p(x)$ tiene un único máximo en el punto $x = a$ (en donde $p'(x) = 0$), y toma el valor $p(a) = 1/(\sigma\sqrt{2\pi})$; que se verifica $p(x) \neq 0$ para todo x real, y que $\lim_{x \rightarrow -\infty} p(x) = \lim_{x \rightarrow +\infty} p(x) = 0$. La recta $x = a$ es eje de simetría de la curva $y = p(x)$. Se puede ver, que en los dos valores $x = a \pm \sigma$, el gráfico de $p(x)$ presenta puntos de inflexión. El gráfico de la función $p(x)$ se representa en la figura 3.5.

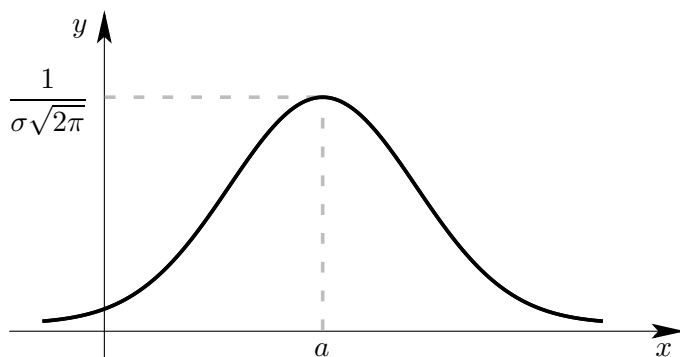


Figura 3.5: Gráfico de la densidad normal con parámetros (a, σ)

Si variamos el valor de a , manteniendo σ constante, el gráfico de la función $p(x)$ se traslada, sin cambiar de forma. Si fijamos a y tomamos dos valores de σ , por ejemplo, $\sigma_1 < \sigma_2$, los gráficos de las densidades normales $p_1(x)$ y $p_2(x)$ con parámetros (a, σ_1) y (a, σ_2) respectivamente, presentan máximo en el mismo punto $x = a$, con valores máximos diferentes $1/(\sigma_1\sqrt{2\pi}) > 1/(\sigma_2\sqrt{2\pi})$. Teniendo en cuenta, que el área bajo cada uno de los gráficos de las densidades $p_1(x)$ y $p_2(x)$ es igual a 1, (por (3.6)), estos gráficos se representan como en la figura 3.6.

La distribución $F(x)$ de la variable aleatoria considerada, es

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(t-a)^2/2\sigma^2} dt,$$

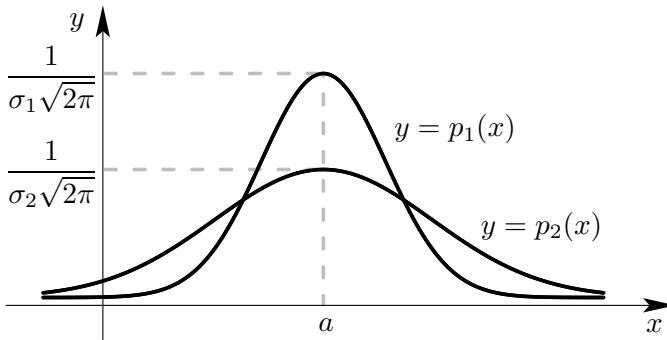


Figura 3.6: Comparación de densidades normales con $\sigma_1 < \sigma_2$

en vista de (3.5).

El caso particular en el que $a = 0$ y $\sigma = 1$ corresponde a la denominada *distribución normal estándar*. Una variable aleatoria tiene distribución normal estándar, si tiene densidad dada por

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (3.13)$$

La distribución correspondiente a esta variable aleatoria se denota mediante $\Phi(x)$. De esta forma, la distribución normal estándar se define por la fórmula

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Las funciones $\varphi(x)$ y $\Phi(x)$, relacionadas por la igualdad $\Phi'(x) = \varphi(x)$, fueron introducidas en el capítulo 2 y están tabuladas. Se puede demostrar que la densidad de una variable aleatoria con distribución t con n grados de libertad definida en (3.10), tiene como límite, cuando $n \rightarrow \infty$, a la densidad normal $\varphi(x)$, de forma que para valores grandes de n , la densidad (3.10) es similar a $\varphi(x)$.

De (3.12) y (3.13) se obtienen en forma directa las fórmulas

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right), \quad p(x) = \frac{1}{\sigma} \varphi\left(\frac{x-a}{\sigma}\right), \quad (3.14)$$

válidas para todo x real.

Problema. Sean $\alpha < \beta$ reales y X una variable aleatoria con distribución normal con parámetros (a, σ) . Calcular $\mathbf{P}(\alpha \leq X \leq \beta)$.

Solución. Según las fórmulas (3.7) y (3.14), se tiene

$$\mathbf{P}(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x)dx = F(\beta) - F(\alpha) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \quad (3.15)$$

El último término en (3.15) se puede calcular dados los valores de α , β , a y σ , utilizando los valores de la tabla de la función $\Phi(x)$, presentada al final del libro.

Como casos particulares, si X es una variable aleatoria con distribución normal con parámetros (a, σ) , de la fórmula (3.15) se obtienen las siguientes probabilidades, que aparecen usualmente en aplicaciones estadísticas:

$$\begin{aligned} \mathbf{P}(a - \sigma \leq X \leq a + \sigma) &= 0,68 \\ \mathbf{P}(a - 1,96\sigma \leq X \leq a + 1,96\sigma) &= 0,95 \\ \mathbf{P}(a - 3\sigma \leq X \leq a + 3\sigma) &= 0,997 \end{aligned}$$

En el gráfico 3.7, el área de la figura delimitada por el gráfico de la función

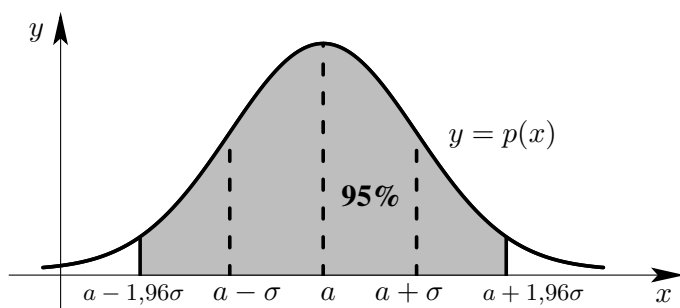


Figura 3.7: Gráfico de la densidad normal $p(x)$ con parámetros (a, σ) . El área sombreada es el 95 % del área total

$p(x)$, el eje Ox , y las rectas $x = a - 1,96\sigma$ y $x = a + 1,96\sigma$, representa el 95 % del total del área entre la gráfica y el eje Ox (que es igual a 1).

Ejemplo 3.12. Decimos que una variable aleatoria X tiene *distribución Gama* con parámetros (α, λ) , donde $\alpha > 0$ y $\lambda > 0$, si tiene densidad dada por

$$p(x) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases} \quad (3.16)$$

Aquí $\Gamma(\lambda)$ es la función Gama definida en (3.11). De la definición de esta función, es sencillo obtener la igualdad $\Gamma(n) = (n-1)\Gamma(n-1)$, válida para cualquier natural $n \geq 2$. Teniendo en cuenta que $\Gamma(1) = 1$, obtenemos la fórmula $\Gamma(n) = (n-1)!$, válida para cualquier natural $n \geq 1$. Por ésto, si ponemos $\lambda = 1$ en la densidad (3.16), obtenemos la densidad exponencial definida en (3.9), que resulta ser un caso particular de la densidad gama.

Otro caso particular importante es el que se obtiene cuando $\alpha = 1/2$ y $\lambda = n/2$, que corresponde a la denominada *distribución χ^2* con n grados de libertad⁵. Junto con la distribución t , definida en (3.10), la distribución χ^2 , que tiene densidad dada por

$$p(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases}$$

juega un rol muy importante en estadística. Veamos el gráfico de la densidad χ^2 con $n = 1, n = 2$ y $n = 4$ grados de libertad en la figura 3.8.

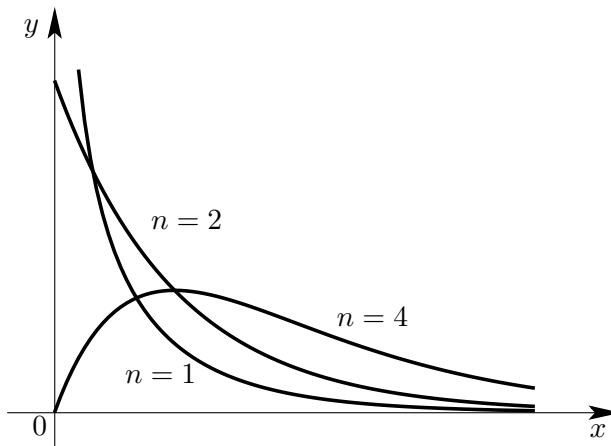


Figura 3.8: Gráficos de densidades χ^2 para $n = 1, 2$ y 4 grados de libertad.

Hasta el momento hemos considerado únicamente variables aleatorias con distribuciones de dos tipos: discretas y absolutamente continuas. Estos tipos no agotan todas las posibilidades. Por ejemplo, una variable aleatoria con función de distribución que tenga derivada $(\pi(1+x^2))^{-1}$ para $x \leq 0$,

⁵La distribución χ^2 , se lee *ji cuadrado*

y que tome en la semirrecta positiva únicamente los valores 1 y 2 con probabilidad $1/4$ en cada uno, tiene una distribución que no resulta ser ni discreta, ni absolutamente continua. Existe además una tercer clase de distribuciones, denominadas *distribuciones singulares*. Una función de distribución singular $F(x)$ es continua para todo x , y verifica $F'(x) = 0$ casi seguramente, con respecto a la medida de Lebesgue⁶.

3.3. Vectores aleatorios y variables aleatorias independientes.

Sean X_1, \dots, X_n variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. El vector $X = (X_1, \dots, X_n)$ se denomina *vector aleatorio*, o también *variable aleatoria n -dimensional*. Este vector toma valores en \mathbb{R}^n , el espacio euclideo de dimensión n . En el caso particular $n = 1$, que llamamos *unidimensional*, obtenemos una variable aleatoria. Como el conjunto $\{\omega: X_k(\omega) \leq x_k\}$ es un suceso (es decir, pertenece a \mathcal{A}) para cada $k = 1, \dots, n$ y reales x_1, \dots, x_k arbitrarios, tenemos que

$$\bigcap_{i=1}^n \{\omega: X_k(\omega) \leq x_k\} \in \mathcal{A},$$

y se puede definir la función real de n variables

$$F(x_1, \dots, x_n) = \mathbf{P}\left(\bigcap_{i=1}^n \{\omega: X_k(\omega) \leq x_k\}\right),$$

que se denomina *función de distribución n -dimensional* del vector aleatorio X . La probabilidad recién considerada se designa también $\mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$. Luego, la definición dada, es la identidad

$$F(x_1, \dots, x_n) = \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n). \quad (3.17)$$

Para $n = 1$ la definición en (3.17) coincide con la definición de distribución de una variable aleatoria (3.3).

Una función de distribución $F(x_1, \dots, x_n)$ cumple las siguientes propiedades.

⁶Una distribución de este tipo es la dada por la *función de Cantor*. Los detalles pueden verse, por ejemplo, en [1].

Propiedad 1. Se verifica $0 \leq F(x_1, \dots, x_n) \leq 1$, para todo x_1, \dots, x_n .

Propiedad 2. La función $F(x_1, \dots, x_n)$ es no decreciente en cada uno de sus argumentos.

Propiedad 3. Se tiene

$$\begin{aligned} \lim_{x_n \rightarrow +\infty} F(x_1, \dots, x_n) &= \mathbf{P}(X_1 \leq x_1, \dots, X_{n-1} \leq x_{n-1}), \\ \lim_{x_n \rightarrow -\infty} F(x_1, \dots, x_n) &= 0. \end{aligned}$$

La propiedad 1 es evidente. Las demostraciones de las propiedades 2 y 3 son análogas a las correspondientes al caso de variables aleatorias ($n = 1$).

Como en el caso unidimensional, los dos tipos más importantes de distribuciones n -dimensionales son las discretas y las absolutamente continuas.

Decimos que un vector aleatorio $X = (X_1, \dots, X_n)$ tiene *distribución discreta*, si existe un conjunto B de puntos del espacio euclideo \mathbb{R}^n , finito o numerable, tal que se verifica $\mathbf{P}(X \in B) = 1$. El vector aleatorio X tiene *distribución absolutamente continua*, cuando su función de distribución $F(x_1, \dots, x_n)$ puede representarse de la forma

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(u_1, \dots, u_n) du_1 \dots du_n \quad (3.18)$$

para reales x_1, \dots, x_n arbitrarios, donde la función $p(u_1, \dots, u_n)$ es no negativa e integrable, y se denomina *densidad del vector aleatorio X* . Como en el caso unidimensional, tiene lugar la identidad

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(u_1, \dots, u_n) du_1 \dots du_n = 1. \quad (3.19)$$

Para $n = 1$ ambas definiciones se reducen a las respectivas definiciones de la sección 3.2.

Un ejemplo importante de distribución n -dimensional discreta es la distribución multinomial.

Ejemplo 3.13. Consideremos n números positivos p_1, \dots, p_n , que verifican la condición $p_1 + \cdots + p_n = 1$, y un natural $N \geq 2$. Decimos que el

vector aleatorio $X = (X_1, \dots, X_n)$ tiene *distribución multinomial*, con parámetros (N, p_1, \dots, p_n) , si se verifica

$$\mathbf{P}(X_1 = m_1, \dots, X_n = m_n) = \frac{N!}{m_1! \cdots m_n!} p_1^{m_1} \cdots p_n^{m_n},$$

para todos los posibles $m_k = 0, \dots, N$ ($k = 1, \dots, n$), que verifican $m_1 + \cdots + m_n = N$. En el caso $N = 2$ se obtiene la distribución binomial con parámetros (n, p_1) , introducida en la sección 2.1, y considerada en el ejemplo 3.5.

Veamos ahora ejemplos de distribuciones n -dimensionales absolutamente continuas.

Ejemplo 3.14. Decimos que el vector aleatorio $X = (X_1, \dots, X_n)$ tiene *distribución uniforme* en una región \mathbf{D} del espacio \mathbb{R}^n , si tiene densidad dada por

$$p(x_1, \dots, x_n) = \begin{cases} c, & \text{si } (x_1, \dots, x_n) \in \mathbf{D}, \\ 0, & \text{en caso contrario.} \end{cases}$$

donde c es una constante positiva, que se determina mediante la condición (3.19).

Ejemplo 3.15. Decimos que el vector aleatorio $X = (X_1, \dots, X_n)$ tiene *distribución normal n -dimensional*, si tiene densidad dada por

$$p(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} e^{-\frac{1}{2}(x-a)B^{-1}(x-a)} \quad (3.20)$$

donde $x = (x_1, \dots, x_n)$ y $a = (a_1, \dots, a_n)$ son vectores fila de números reales; B es una matriz de dimensión $n \times n$, definida positiva⁷, no singular y simétrica, B^{-1} es la matriz inversa de la matriz B , y x' denota el vector traspuesto de x .

En el caso $n = 1$, la densidad dada en (3.20) se reduce a la fórmula de la densidad normal (3.12), donde el “vector” a es el número a , y la matriz de dimensión 1×1 es $B = [\sigma^2]$. Consideraremos nuevamente la distribución normal n -dimensional en el capítulo 4.

Examinemos ahora el caso $n = 2$, es decir, consideremos un vector aleatorio (X, Y) con *distribución normal bidimensional*⁸. No es difícil de

⁷Una matriz $B = (b_{ij})$ de dimensión $n \times n$ es *definida positiva*, si para todo vector fila $x = (x_1, \dots, x_n)$ no nulo, se verifica $xBx' = \sum_{i,j} x_i b_{ij} x_j > 0$.

⁸Decimos *bidimensional* en vez de 2-dimensional.

verificar que la matriz

$$B = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}, \quad (3.21)$$

verifica las tres condiciones indicadas, si $\sigma_1 > 0$, $\sigma_2 > 0$, y $-1 < \rho < 1$. Si $a = (a_1, a_2)$, y sustituimos la matriz B dada en (3.21) en la fórmula (3.20), obtenemos, que la densidad del vector (X, Y) está dada por

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\left(\frac{x-a_1}{\sigma_1} \right)^2 + \left(\frac{y-a_2}{\sigma_2} \right)^2 - 2\rho \frac{(x-a_1)}{\sigma_1} \frac{(y-a_2)}{\sigma_2} \right] \right\}. \quad (3.22)$$

Demostremos que la variable aleatoria X tiene distribución normal con parámetros (a_1, σ_1) , y que la variable aleatoria Y tiene distribución normal con parámetros (a_2, σ_2) . Para x real, tenemos

$$\begin{aligned} \mathbf{P}(X \leq x) &= \mathbf{P}(X \leq x, Y < \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} p(u, v) du dv \\ &= \int_{-\infty}^x \left(\int_{-\infty}^{\infty} p(u, v) dv \right) du. \end{aligned} \quad (3.23)$$

Para calcular la integral con respecto de v , introducimos el cambio de variable $t = ((v - a_2)/\sigma_2 - \rho(u - a_1)/\sigma_1)/\sqrt{1 - \rho^2}$. Calculando t^2 y sustituyendo en el exponente en (3.22), obtenemos

$$\begin{aligned} \int_{-\infty}^{\infty} p(u, v) dv &= \frac{1}{2\pi\sigma_1} \int_{-\infty}^{\infty} e^{-(t^2/2 + (u-a_1)^2/(2\sigma_1^2))} dt \\ &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(u-a_1)^2/(2\sigma_1^2)}, \end{aligned}$$

donde utilizamos que $\int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$. Sustituyendo esta expresión en (3.23), resulta

$$\mathbf{P}(X \leq x) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^x e^{-(u-a_1)^2/(2\sigma_1^2)} du,$$

es decir, la variable aleatoria X tiene distribución normal con parámetros (a_1, σ_1) . La afirmación relativa a Y se demuestra en forma análoga.

Introducimos ahora la noción de *independencia* para variables aleatorias.

Definición 3.1. Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, en el cual están definidas las variables aleatorias X_1, \dots, X_n . Designamos mediante $F_k(x)$ a la distribución de la variable aleatoria X_k ($k = 1, \dots, n$), y mediante $F(x_1, \dots, x_n)$ a la distribución del vector aleatorio (X_1, \dots, X_n) . Decimos que X_1, \dots, X_n son variables aleatorias mutuamente independientes, o más brevemente variables aleatorias independientes, cuando se verifica

$$F(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n) \quad (3.24)$$

para reales x_1, \dots, x_n arbitrarios.

Es fácil de ver que las variables aleatorias X_1, \dots, X_n son independientes si y solo si, para reales x_1, \dots, x_n arbitrarios, son independientes los sucesos $\{\omega: X_1(\omega) \leq x_1\}, \dots, \{\omega: X_n(\omega) \leq x_n\}$ (ver (1.12), definición de independencia de sucesos⁹).

Si consideramos variables aleatorias discretas o absolutamente continuas, podemos formular la condición (3.24) que define su independencia, en términos de las probabilidades de los valores que toman las variables aleatorias discretas, o en términos de densidades en el caso absolutamente continuo. Para ésto, utilizamos el resultado siguiente.

Lema 3.2. Consideremos dos variables aleatorias X, Y independientes. Para $a < b, c < d$, se verifica

$$\mathbf{P}(a < X \leq b, c < Y \leq d) = \mathbf{P}(a < X \leq b) \mathbf{P}(c < Y \leq d).$$

Demostración. Sea $F(x, y)$ la función de distribución del vector (X, Y) . Tenemos (ver figura 3.9)

$$\begin{aligned} F(b, d) &= \mathbf{P}(X \leq b, Y \leq d) = \mathbf{P}(X \leq a, Y \leq d) + \mathbf{P}(X \leq b, Y \leq c) \\ &\quad - \mathbf{P}(X \leq a, Y \leq c) + \mathbf{P}(a < X \leq b, c < Y \leq d). \end{aligned}$$

Aplicando la fórmula (3.24), obtenemos

⁹Es posible demostrar que la validez de la condición (3.24) para x_1, \dots, x_n arbitrarios es equivalente a la independencia de los n sucesos $\{\omega: X_k(\omega) \in B_k\}$ ($k = 1, \dots, n$), donde B_1, \dots, B_n son conjuntos de Borel arbitrarios de la recta real.

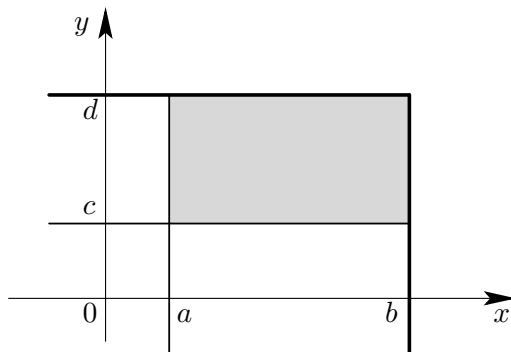


Figura 3.9: El conjunto $\{a < x \leq b, c < y \leq d\}$

$$\begin{aligned}
 \mathbf{P}(a < X \leq b, c < Y \leq d) &= \mathbf{P}(X \leq b) \mathbf{P}(Y \leq d) - \mathbf{P}(X \leq a) \mathbf{P}(Y \leq d) \\
 &\quad - \mathbf{P}(X \leq b) \mathbf{P}(Y \leq c) + \mathbf{P}(X \leq a) \mathbf{P}(Y \leq c) \\
 &= \mathbf{P}(X \leq b) (\mathbf{P}(Y \leq d) - \mathbf{P}(Y \leq c)) \\
 &\quad - \mathbf{P}(X \leq a) (\mathbf{P}(Y \leq d) - \mathbf{P}(Y \leq c)) \\
 &= \mathbf{P}(a < X \leq b) \mathbf{P}(c < Y \leq d),
 \end{aligned}$$

lo que concluye la demostración. \square

Proposición 3.1. *Consideremos dos variables aleatorias X, Y con distribución discreta. Sean x_1, x_2, \dots los valores que toma la variable X ; y_1, y_2, \dots los valores que toma la variable Y . Las variables aleatorias X e Y son independientes, si y solo si se verifica*

$$\mathbf{P}(X = x_k, Y = y_j) = \mathbf{P}(X = x_k) \mathbf{P}(Y = y_j), \quad k, j = 1, 2, \dots \quad (3.25)$$

Demostración. Supongamos primero que las variables aleatorias X e Y son independientes. Para cada k y cada n naturales, consideramos el suceso

$$\mathbf{A}_{k,n} = \{\omega : x_k - 1/n < X(\omega) \leq x_k\}.$$

Tenemos $\mathbf{A}_{k,1} \supset \mathbf{A}_{k,2} \supset \dots$, y además $\bigcap_{n=1}^{\infty} \mathbf{A}_{k,n} = \{\omega : X(\omega) = x_k\}$. Por la propiedad 8 en la sección 1.3, obtenemos $\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{A}_{k,n}) = \mathbf{P}(X = x_k)$. En forma análoga, obtenemos $\lim_{n \rightarrow \infty} \mathbf{P}(y_j - 1/n < Y \leq y_j) = \mathbf{P}(Y = y_j)$ para j arbitrario. Con la misma argumentación, obtenemos que

$$\lim_{n \rightarrow \infty} \mathbf{P}(x_k - 1/n < X \leq x_k, y_j - 1/n < Y \leq y_j) = \mathbf{P}(Y = y_j, X = x_k) \quad (3.26)$$

para j, k arbitrarios. Aplicando el lema 3.2, se tiene $\mathbf{P}(x_k - 1/n < X \leq x_k, y_j - 1/n < Y \leq y_j) = \mathbf{P}(x_k - 1/n < X \leq x_k) \mathbf{P}(y_j - 1/n < Y \leq y_j)$, que sustituido en (3.26), da la igualdad (3.25).

Consideremos ahora el recíproco. Supongamos que las variables aleatorias X e Y verifican la condición (3.25). Consideremos el vector aleatorio (X, Y) y su función de distribución $F(x, y) = \mathbf{P}(X \leq x, Y \leq y)$. Para demostrar la independencia de las variables aleatorias X e Y , tenemos que demostrar $F(x, y) = \mathbf{P}(X \leq x) \mathbf{P}(Y \leq y)$, para x e y arbitrarios. Es claro que $F(x, y) = \sum_{k,j} \mathbf{P}(X = x_k, Y = y_j)$, donde la suma afecta a aquellos valores de k y j , para los cuales $x_k \leq x$ e $y_j \leq y$ (ver figura 3.10). Aplicando (3.25), obtenemos

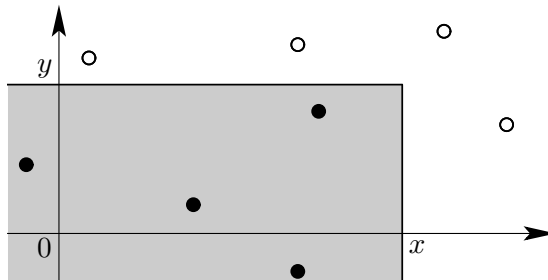


Figura 3.10: En negro se indican los puntos incluidos en la suma.

$$\begin{aligned} F(x, y) &= \sum_{k,j} \mathbf{P}(X = x_k) \mathbf{P}(Y = y_j) \\ &= \sum_{k: x_k \leq x} \mathbf{P}(X = x_k) \sum_{j: y_j \leq y} \mathbf{P}(Y = y_j) = \mathbf{P}(X \leq x) \mathbf{P}(Y \leq y), \end{aligned}$$

concluyendo la demostración. \square

Utilizando un método similar, se obtiene la siguiente generalización de la proposición 3.1.

Proposición 3.2. *Consideremos n variables aleatorias X_1, \dots, X_n con distribución discreta, siendo x_{k1}, x_{k2}, \dots los valores que toma cada variable aleatoria X_k , para $k = 1, \dots, n$. Las variables aleatorias dadas son independientes, si y solo si se verifica*

$$\mathbf{P}(X_1 = x_{1m_1}, \dots, X_n = x_{nm_n}) = \prod_{k=1}^n \mathbf{P}(X_k = x_{km_k}),$$

para naturales m_1, \dots, m_n no nulos y arbitrarios.

El resultado que sigue es el análogo de la proposición 3.2 para variables aleatorias absolutamente continuas.

Proposición 3.3. *Consideremos un vector aleatorio $X = (X_1, \dots, X_n)$ absolutamente continuo, y sea $p_k(x)$ la densidad de la variable aleatoria X_k ($k = 1, \dots, n$). Las variables aleatorias X_1, \dots, X_n son independientes, si y solo si se verifica*

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n) \quad (3.27)$$

*casi seguramente*¹⁰.

Demostración. Supongamos en primer lugar que las variables aleatorias X_1, \dots, X_n son independientes. Luego, la función de distribución se escribe de dos formas: como producto de las distribuciones $F_k(x_k)$ ($k = 1, \dots, n$), y como integral múltiple de la densidad $p(u_1, \dots, u_n)$, es decir

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(u_1, \dots, u_n) du_1 \cdots du_n = F_1(x_1) \cdots F_n(x_n).$$

Derivando n veces en ambos miembros de esta identidad, primero respecto de x_1 , luego respecto de x_2, \dots , y finalmente respecto de x_n , obtenemos la fórmula (3.27).

Supongamos ahora que se cumple (3.27). Integramos n veces en ambos lados de esta igualdad, primero con respecto de x_1 en el intervalo $(-\infty, y_1)$, \dots , y por último con respecto de x_n en el intervalo $(-\infty, y_n)$. Teniendo en cuenta (3.18), obtenemos

$$F(y_1, \dots, y_n) = F_1(y_1) \cdots F_n(y_n).$$

que se cumple para y_1, \dots, y_n arbitrarios. Como $F_k(y) = \int_{-\infty}^y p_k(x) dx = \mathbf{P}(X_k \leq y)$, se verifica la definición de independencia. \square

Ejemplo 3.16. Consideremos nuevamente un vector aleatorio (X, Y) con distribución normal bidimensional, con densidad $p(x, y)$ dada en (3.22). Hemos demostrado, que la variable aleatoria X tiene distribución normal con parámetros (a_1, σ_1) , y que la variable aleatoria Y tiene distribución

¹⁰Es decir, en todos los puntos, exceptuando un conjunto de medida de Lebesgue nula.

normal con parámetros (a_2, σ_2) . Si $\rho = 0$, la densidad en (3.22) se reduce a

$$p(x, y) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-(x-a_1)^2/(2\sigma_1^2)} \times \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-(y-a_2)^2/2\sigma_2^2} = p_1(x)p_2(y),$$

donde $p_1(x)$ es la densidad de la variable aleatoria X , y $p_2(y)$ la densidad de Y . Como consecuencia de la proposición 3.3, en el caso $\rho = 0$, las variables aleatorias X e Y son independientes.

Ejemplo 3.17. Consideremos un vector aleatorio (X_1, \dots, X_n) con distribución normal n -dimensional, con densidad $p(x)$ dada en la fórmula (3.20). Supongamos que la matriz B es diagonal, y está dada por

$$B = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}. \quad (3.28)$$

Si $\sigma_k > 0$ ($k = 1, \dots, n$) la matriz dada verifica las tres condiciones indicadas en la definición (ser definida positiva, simétrica, y no singular). Sustituyendo la matriz B en la fórmula (3.20), obtenemos la densidad del vector considerado, que es

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma_1 \cdots \sigma_n} e^{-\sum_{k=1}^n (x_k - a_k)^2 / (2\sigma_k^2)}. \quad (3.29)$$

Como ocurre en el caso $n = 2$, se verifica que cada variable aleatoria X_k tiene distribución normal, con parámetros (a_k, σ_k) , es decir, tiene densidad $p_k(x) = (\sigma_k \sqrt{2\pi})^{-1} e^{-(x-a_k)^2/(2\sigma_k^2)}$, para $k = 1, \dots, n$. Mas aún, factorizando la fórmula (3.29), tenemos

$$p(x_1, \dots, x_n) = \prod_{k=1}^n \frac{1}{\sigma_k \sqrt{2\pi}} e^{-(x_k - a_k)^2 / (2\sigma_k^2)} = \prod_{k=1}^n p_k(x_k). \quad (3.30)$$

Aplicando la proposición 3.3 se obtiene la independencia mutua de las variables aleatorias X_1, \dots, X_n .

Recíprocamente, si las variables aleatorias X_1, \dots, X_n son independientes, y cada una de las variables X_k tiene distribución normal con parámetros (a_k, σ_k) ($k = 1, \dots, n$), aplicando la proposición 3.3, obtenemos que la densidad $p(x_1, \dots, x_n)$ del vector $X = (X_1, \dots, X_n)$ verifica la fórmula (3.30), y en consecuencia, que el vector X tiene distribución normal n -dimensional, con densidad dada en (3.20), y matriz B dada en (3.28).

3.4. Distribución de la suma de variables aleatorias independientes

Demostremos la siguiente proposición.

Proposición 3.4. Sean X_1 y X_2 variables aleatorias independientes, con densidades $p_1(x)$ y $p_2(x)$ respectivamente. La suma $X_1 + X_2$ tiene densidad dada por

$$p(x) = \int_{-\infty}^{\infty} p_1(u)p_2(x - u)du. \quad (3.31)$$

Demostración. Sea $p(u, v)$ la densidad del vector (X_1, X_2) . Como las variables X_1 y X_2 son independientes, en vista de la proposición 3.3, tenemos $p(u, v) = p_1(u)p_2(v)$. Consideremos el valor de la función de distribución de la suma $X_1 + X_2$ en un punto x arbitrario:

$$\mathbf{P}(X_1 + X_2 \leq x) = \mathbf{P}(X_1 + X_2 \in \mathbf{D}) = \int \int_{\mathbf{D}} p(u, v)dudv,$$

donde $\mathbf{D} = \{(u, v) : u + v \leq x\}$ es la región bajo la recta de ecuación $u + v = x$ (ver figura 3.11). De aquí obtenemos

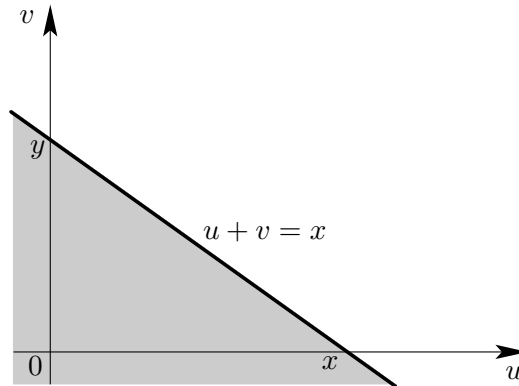


Figura 3.11: Región $\mathbf{D} = \{(u, v) : u + v \leq x\}$

$$\mathbf{P}(X_1 + X_2 \leq x) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x-u} p_2(v)dv \right) p_1(u)du =$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^x p_2(y-u)dy \right) p_1(u)du \\ &= \int_{-\infty}^x \left(\int_{-\infty}^{\infty} p_1(u)p_2(y-u)du \right) dy \end{aligned}$$

para x arbitrario. En consecuencia la distribución de la suma $X_1 + X_2$ es absolutamente continua, y tiene densidad $p(x) = \int_{-\infty}^{\infty} p_1(u)p_2(x-u)du$. \square

Observemos, que intercambiando los roles de X_1 y X_2 , obtenemos también la fórmula

$$p(x) = \int_{-\infty}^{\infty} p_1(x-u)p_2(u)du \tag{3.32}$$

Fórmulas análogas tienen lugar para la distribución de una suma de dos variables aleatorias independientes, que tienen distribuciones discretas. En este caso, en lugar de integrales aparecen sumas. Es posible obtener fórmulas mas generales (que incluyan ambos tipos de distribuciones) para la función de distribución $F(x)$ de una suma de variables aleatorias independientes X_1 y X_2 , con funciones de distribución $F_1(x)$ y $F_2(x)$. Mas precisamente, se trata de la fórmula

$$F(x) = \int_{-\infty}^{+\infty} F_1(x-y)dF_2(y) = \int_{-\infty}^{+\infty} F_2(x-y)dF_1(y), \tag{3.33}$$

en las cuales se utiliza la integral de Stieltjes. Las integrales anteriores y las que figuran en las identidades (3.31) y (3.32), se denominan *convolución* o *composición* de las distribuciones.

Apliquemos la proposición 3.4 para demostrar que la suma de variables aleatorias independientes, cada una de las cuales tiene distribución normal, tiene distribución normal.

Ejemplo 3.18. Consideremos entonces dos variables aleatorias independientes X_1 y X_2 , tales que X_k tiene densidad $p_k(x) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-(x-a_k)^2/(2\sigma_k^2)}$ ($k = 1, 2$). Demostremos que $X_1 + X_2$ tiene distribución normal con parámetros $(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$, es decir, que tiene densidad dada por

$$p(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-(x-a_1-a_2)^2/(2(\sigma_1^2+\sigma_2^2))}.$$

Demostremos primero esta afirmación en el caso particular $a_1 = a_2 = 0$. Aplicando la fórmula de convolución (3.32), obtenemos

$$p(x) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-(x-u)^2/(2\sigma_1^2)-u^2/(2\sigma_2^2)} du.$$

Introduciendo una nueva variable de integración, según la fórmula $v = u \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} - \frac{\sigma_2 x}{\sigma_1 \sqrt{\sigma_1^2 + \sigma_2^2}}$, resulta

$$\begin{aligned} p(x) &= \frac{1}{2\pi \sqrt{\sigma_1^2 + \sigma_2^2}} \int_{-\infty}^{\infty} e^{-v^2/2 - x^2/(2(\sigma_1^2 + \sigma_2^2))} dv \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-x^2/(2(\sigma_1^2 + \sigma_2^2))}. \end{aligned}$$

De esta forma, la densidad $p(x)$ de la suma $X_1 + X_2$ es normal con parámetros $(0, \sqrt{\sigma_1^2 + \sigma_2^2})$.

La demostración cuando los parámetros a_1 y a_2 son arbitrarios se reduce al caso anterior ($a_1 = a_2 = 0$), mediante la consideración de nuevas variables aleatorias $Y_k = X_k - a_k$ ($k = 1, 2$). Estas variables son independientes, por serlo las variables X_1 y X_2 , tienen función de distribución

$$\mathbf{P}(Y_k \leq x) = \mathbf{P}(X_k \leq x + a_k) = F_k(x + a_k) = \int_{-\infty}^{x+a_k} p_k(u) du,$$

y densidad dada por $\frac{d}{dx} P(Y_k \leq x) = p_k(x + a_k) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-x^2/(2\sigma_k^2)}$ ($k = 1, 2$), que es la densidad normal con parámetros $(0, \sigma_k)$. Según vimos, la suma $Y_1 + Y_2$ tiene distribución normal con parámetros $(0, \sqrt{\sigma_1^2 + \sigma_2^2})$. En consecuencia, la variable aleatoria $X_1 + X_2 = Y_1 + Y_2 + a_1 + a_2$ tiene distribución normal con parámetros $(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

En el capítulo 7 daremos otra demostración de este hecho, basado en el método de las funciones características, que utiliza cálculos más sencillos.

3.5. Ejercicios

1. La variable aleatoria X tiene distribución discreta y toma los valores 0, 1, 2 y 4, con probabilidades $1/2, 1/4, 1/8$ y $1/8$. Hallar la función de distribución $F(x)$ de esta variable aleatoria, y dibujar el gráfico de $y = F(x)$.
2. La variable aleatoria X tiene distribución uniforme en el intervalo $(0, 2)$. Hallar la función de distribución y su densidad. Dibujar los gráficos correspondientes.
3. En el ejercicio 2, calcular $\mathbf{P}(0,5 \leq X \leq 1,5)$.

4. Consideremos una variable aleatoria X , con distribución normal, con parámetros $(2, 3/2)$. Calcular las siguientes probabilidades: (a) $\mathbf{P}(X \geq 3)$; (b) $\mathbf{P}(1 \leq X \leq 4)$; (c) $\mathbf{P}(|X - 2,5| \geq 0,5)$; (d) $\mathbf{P}((X - 1)^2 \leq 4)$.

5. La variable aleatoria X tiene densidad $p(x) = ce^{-|x|}$. (a) Calcular $\mathbf{P}(X \leq 0)$; (b) hallar el valor de la constante c ; y (c) calcular $\mathbf{P}(0 \leq X \leq 1)$.

6. La variable aleatoria X tiene distribución normal con parámetros $(3, 4)$. (a) Calcular las probabilidades $\mathbf{P}(X < 0)$ y $\mathbf{P}(-9 < X < 1)$. (b) Hallar el valor de x que cumple la condición $\mathbf{P}(X > x) = 0,01$.

7. La variable aleatoria X tiene densidad dada por

$$p(x) = \begin{cases} cx^2, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{si } x < 0 \text{ ó } x > 1 \end{cases}$$

(a) Hallar el valor de c . (b) Hallar la función de distribución de X , y (c) calcular la probabilidad $\mathbf{P}(0,1 < X < 0,4)$.

8. Una variable aleatoria con función de distribución

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

donde $\sigma > 0$, se denomina *distribución de Rayleigh*. Hallar la densidad de esta variable aleatoria, y calcular la probabilidad $\mathbf{P}(0 \leq X \leq \sigma)$.

9. Verificar, que la función dada por

$$p(x) = \begin{cases} \frac{a}{b} \left(\frac{b}{x}\right)^{a+1}, & x > b, \\ 0, & x < b. \end{cases}$$

donde a y b son constantes positivas, es la densidad correspondiente una variable aleatoria, y calcular su función de distribución correspondiente. (Esta distribución se denomina *distribución de Pareto*.)

10. El vector aleatorio (X, Y) tiene distribución discreta, dada en la siguiente tabla:

$x \setminus y$	0	1	2	3
0	0,2	0	0,1	0
1	0,1	0,2	0,1	0
2	0	0,1	0,1	0,1

donde se indican las probabilidades $\mathbf{P}(X = x, Y = y)$, de forma que, por ejemplo, $\mathbf{P}(X = 2, Y = 3) = 0,1$. Hallar $\mathbf{P}(X = 0)$, $\mathbf{P}(X = 1)$, $\mathbf{P}(X = 2)$, y la función de distribución de la variable aleatoria X .

11. En las condiciones del ejercicio 10, calcular $\mathbf{P}(Y = 0)$, $\mathbf{P}(Y = 1)$, $\mathbf{P}(Y = 2)$, $\mathbf{P}(Y = 3)$ y hallar la función de distribución de la variable aleatoria Y .

12. El vector aleatorio (X, Y) tiene densidad dada por

$$p(x, y) = \begin{cases} cxy^2, & 0 \leq x \leq 2, 0 \leq y \leq 1, \\ 0, & \text{en caso contrario.} \end{cases}$$

donde c es una constante. Hallar: (a) el valor de c ; (b) la función de distribución de la variable aleatoria X ; (c) la función de distribución de la variable aleatoria Y ; (d) las densidades de las variables aleatorias X e Y .

13. ¿Son independientes las variables aleatorias X e Y del ejercicio 12? ¿Son independientes las variables aleatorias X e Y del ejercicio 10?

14. En las condiciones del ejercicio 12, calcular las probabilidades $\mathbf{P}(X \leq 1)$, $\mathbf{P}(0,5 \leq Y \leq 1)$ y $\mathbf{P}(X \leq 1, 0,5 \leq Y \leq 1)$.

15. El vector aleatorio (X, Y) tiene densidad dada por

$$p(x) = \begin{cases} cxe^{-y}, & 0 \leq x \leq 1, 0 < y < \infty, \\ 0, & \text{en caso contrario.} \end{cases}$$

donde c es una constante. (a) Hallar el valor de c . (b) ¿Resultan independientes las variables aleatorias X e Y ?

16. La duración en horas de un cierto tipo de lámpara es una variable aleatoria que tiene densidad dada por

$$p(x) = \begin{cases} 0,001e^{-0,001x} & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Se elijen al azar 3 lámparas de una cierta partida. Calcular las probabilidades de que: (a) ni una de las lámparas se tenga que cambiar en el transcurso de las primeras 1000 horas; (b) las tres lámparas tengan que ser cambiadas en el mismo lapso de tiempo.

17. Una urna contiene 6 bolas numeradas del 1 al 6. Se eligen al azar, y sin remplazo, 3 bolas. Sea X la variable aleatoria que indica el mayor de los números obtenido en las bolas elegidas. Hallar la función de distribución de la variable aleatoria X , y calcular $\mathbf{P}(X \geq 5)$.

18. La variable aleatoria X tiene función de distribución $F(x)$ y densidad $p(x)$. Hallar la función de distribución y la densidad de la variable aleatoria $Y = aX + b$, donde $a > 0$ y b son constantes.

19. La variable aleatoria X tiene distribución uniforme en el intervalo $(0, 1)$. Demostrar que la variable aleatoria $Y = 2X - 1$ tiene distribución uniforme en el intervalo $(-1, 1)$.

20. La variable aleatoria X tiene función de distribución $F(x)$ continua y estrictamente creciente. Demostrar que la variable aleatoria $F(X)$ tiene distribución uniforme en el intervalo $(0, 1)$.

21. Consideremos una variable aleatoria Y con distribución uniforme en el intervalo $(0, 1)$, y sea $F(x)$ una función de distribución continua y estrictamente creciente. Demostrar que la variable aleatoria $F^{-1}(Y)$, donde F^{-1} denota la función inversa de la función F , tiene función de distribución $F(x)$.

22. Construir un ejemplo de dos variables aleatorias distintas, que tengan igual función de distribución.

23. La variable aleatoria X tiene función de distribución $F(x)$. Hallar las funciones de distribución de las variables aleatorias $Y = X^3$ y $Z = -X$.

24. Las variables aleatorias X e Y son independientes, y tienen la misma densidad, dada por

$$f(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Demostrar que la variable aleatoria $X + Y$ tiene densidad, dada por

$$g(x) = \begin{cases} xe^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

25. Demostrar la siguiente proposición: Si X e Y son variables aleatorias independientes con distribución de Poisson con parámetros λ_1 y λ_2 respectivamente, entonces $X + Y$ es una variable aleatoria con distribución de Poisson de parámetro $\lambda_1 + \lambda_2$.

26. Sean X e Y variables aleatorias independientes, cada una de las cuales tiene distribución uniforme en el intervalo $(0, 1)$. Hallar la densidad de $X + Y$.

27. Sean X e Y variables aleatorias independientes, con densidades respectivas $p_1(x)$ y $p_2(x)$. Hallar la densidad de la diferencia $X - Y$.

28. (a) Sea X una variable aleatoria con distribución exponencial de parámetro $\alpha > 0$. Demostrar que

$$\mathbf{P}(X > x + y | X > x) = \mathbf{P}(X > y). \quad (3.34)$$

Esta propiedad se denomina *pérdida de memoria*.

(b) Sea $G: [0, \infty) \rightarrow \mathbb{R}$ una función real, que verifica $G(0) = 1$, y cumple la ecuación funcional

$$G(x + y) = G(x)G(y) \text{ para todo } x \geq 0, y \geq 0.$$

(i) Demostrar que $G(x) = G(1)^x$ para todo x racional. (ii) Demostrar que si además la función $G(x)$ es decreciente, entonces existe $\alpha > 0$ tal que $G(x) = e^{-\alpha x}$. (iii) Concluir que una variable aleatoria que cumple la propiedad (3.34) tiene distribución exponencial.

Capítulo 4

Esperanza matemática, varianza, y otros momentos de variables aleatorias

4.1. Esperanza matemática

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega)$, $\omega \in \Omega$. El espacio de probabilidad dado es un espacio medible con una medida, y la variable aleatoria una función medible¹. Podemos en consecuencia introducir la noción de integral. Supongamos entonces que $\int_{\Omega} |X| d\mathbf{P} < \infty$. En este caso decimos que *existe* la *esperanza matemática*, el *valor esperado*, o más brevemente la *esperanza* de la variable aleatoria X , que designamos $\mathbf{E}X$ y definimos mediante la identidad

$$\mathbf{E}X = \int_{\Omega} X d\mathbf{P}. \quad (4.1)$$

Cuando la variable aleatoria X está acotada, es decir $|X(\omega)| \leq C$ para todo $\omega \in \Omega$ y para alguna constante C , existe la esperanza matemática $\mathbf{E}X$ y es válida la desigualdad $|\mathbf{E}X| \leq C$. En efecto, se tiene

$$|\mathbf{E}X| = \left| \int_{\Omega} X d\mathbf{P} \right| \leq \int_{\Omega} |X| d\mathbf{P} \leq C \int_{\Omega} d\mathbf{P} = C \mathbf{P}(\Omega) = C.$$

¹El lector que no se encuentre no familiarizado con la teoría de la medida, puede restringirse a la consideración de variables aleatorias que tengan distribución discreta o absolutamente continua, y tomar, en calidad de definición de esperanza matemática, las fórmulas (4.5) y (4.7) respectivamente, en la página 90.

Para las variables aleatorias no acotadas, la esperanza matemática no necesariamente existe.

Sea X una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. Consideremos una nueva medida de probabilidad \mathbf{P}_X , definida en los conjuntos borelianos B de la recta real, mediante $\mathbf{P}_X(B) = \mathbf{P}(X \in B)$. (Es inmediato verificar que la función \mathbf{P}_X verifica los axiomas de la sección 1.2.) En particular, si $F(x)$ es la distribución de la variable aleatoria X , para los intervalos de la recta real de la forma $(a, b]$ tenemos $\mathbf{P}_X((a, b]) = F(b) - F(a)$.

De esta forma, la variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ genera un nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}, \mathbf{P}_X)$, donde \mathcal{B} es la σ -álgebra de los conjuntos de Borel y \mathbf{P}_X la probabilidad recién definida.

Sea $g(x)$ una función de Borel² definida en la recta real. Del análisis real es conocida la siguiente proposición: si una de las dos integrales $\int_{\Omega} |g(X)| d\mathbf{P}$ ó $\int_{\mathbb{R}} |g| d\mathbf{P}_X$ es finita, entonces es finita la otra, y tiene lugar la identidad

$$\int_{\Omega} g(X) d\mathbf{P} = \int_{\mathbb{R}} g(x) d\mathbf{P}_X. \quad (4.2)$$

Esta proposición se demuestra primero para una variable aleatoria que toma una cantidad finita de valores (una *función simple* en la terminología del análisis real), luego para una variable aleatoria positiva y arbitraria mediante pasaje al límite, y finalmente para una variable aleatoria que toma valores de ambos signos. Nos restringimos aquí a la demostración de la igualdad (4.2) para una variable aleatoria X que toma una cantidad finita de valores x_1, \dots, x_m , con probabilidades p_1, \dots, p_m respectivamente ($p_1 + \dots + p_m = 1$). Calculemos primero la integral en el lado izquierdo de (4.2). Sea $\mathbf{A}_k = \{\omega: X(\omega) = x_k\}$ ($k = 1, \dots, m$). Tenemos

$$\begin{aligned} \int_{\Omega} g(X) d\mathbf{P} &= \sum_{k=1}^m \int_{\mathbf{A}_k} g(X) d\mathbf{P} = \sum_{k=1}^m \int_{\mathbf{A}_k} g(x_k) d\mathbf{P} \\ &= \sum_{k=1}^m g(x_k) \int_{\mathbf{A}_k} d\mathbf{P} = \sum_{k=1}^m g(x_k) \mathbf{P}(\mathbf{A}_k) = \sum_{k=1}^m g(x_k) p_k. \end{aligned}$$

Para el cálculo de la integral en el lado derecho de (4.2), partimos la recta real en los puntos x_1, \dots, x_m , suponiendo que $x_1 < \dots < x_m$, y

²Ver nota en la página 58.

la representamos como la unión de los conjuntos formados por un punto $\{x_1\}, \dots, \{x_m\}$, los intervalos $B_0 = (-\infty, x_1)$ y $B_m = (x_m, \infty)$, y los intervalos finitos $B_k = (x_k, x_{k+1})$ ($k = 1, \dots, m-1$). Entonces

$$\int_{\mathbb{R}} g(x) d\mathbf{P}_X = \sum_{k=1}^m \int_{\{x_k\}} g(x) d\mathbf{P}_X + \sum_{k=0}^m \int_{B_k} g(x) d\mathbf{P}_X.$$

Por la definición de la medida \mathbf{P}_X , tenemos

$$\begin{aligned} \int_{\{x_k\}} g(x) d\mathbf{P}_X &= g(x_k) \mathbf{P}(\mathbf{A}_k), \text{ para } k = 1, \dots, m, \\ \int_{B_k} g(x) d\mathbf{P}_X &= 0, \text{ para } k = 0, \dots, m. \end{aligned}$$

En consecuencia, tenemos

$$\int_{\mathbb{R}} g(x) d\mathbf{P}_X = \sum_{k=1}^m g(x_k) p_k,$$

concluyendo la demostración.

La integral a la derecha en (4.2) también se designa $\int_{-\infty}^{\infty} g(x) dF(x)$, donde $F(x)$ es la función de distribución de la variable aleatoria X . En el caso en que $g(x)$ es continua, se puede demostrar que esta última integral coincide con la integral de Riemann-Stieltjes, para la cual se utiliza la misma notación. La integral de Riemann-Stieltjes se define como el límite de las sumas integrales

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \lim_{\lambda \rightarrow 0} \sum_{k=0}^{n-1} g(\xi_k) (F(x_{k+1}) - F(x_k)),$$

donde se consideran particiones $a = x_0 < x_1 < \dots < x_n = b$ cuya norma es $\lambda = \max_{0 \leq k \leq n-1} (x_{k+1} - x_k)$, siendo $\xi_k \in (x_k, x_{k+1}]$ un punto interior en cada intervalo de la partición.

De la identidad (4.2) se obtiene, utilizando la notación recién introducida, que

$$\mathbf{E} g(X) = \int_{\Omega} g(X) d\mathbf{P} = \int_{\mathbb{R}} g(x) d\mathbf{P}_X = \int_{-\infty}^{\infty} g(x) dF(x), \quad (4.3)$$

donde la esperanza existe, si las integrales anteriores convergen absolutamente. En el caso particular en el que $g(x) = x$ para todo x real, obtenemos

$$\mathbf{E} X = \int_{\Omega} X d\mathbf{P} = \int_{\mathbb{R}} x d\mathbf{P}_X = \int_{-\infty}^{\infty} x dF(x),$$

si las integrales anteriores convergen absolutamente.

Veamos ahora los dos tipos más importantes de distribuciones: discretas y absolutamente continuas.

Esperanza matemática para variables aleatorias con distribución discreta

Consideremos una variable aleatoria X con distribución discreta, que toma los valores x_1, x_2, \dots con probabilidades p_1, p_2, \dots respectivamente. Sea $g(x)$ una función continua. Entonces

$$\mathbf{E} g(X) = \sum_k g(x_k) p_k, \quad (4.4)$$

donde la esperanza existe si la serie anterior es absolutamente convergente, es decir, si $\sum_k |g(x_k)| p_k < \infty$. En el caso particular en el que $g(x) = x$ para todo x real, obtenemos

$$\mathbf{E} X = \sum_k x_k p_k, \quad (4.5)$$

si la serie anterior es absolutamente convergente. La fórmula (4.4) se obtiene a partir de (4.3), y también a partir de (4.1). Si X toma únicamente una cantidad finita de valores x_1, \dots, x_m , existe su esperanza matemática, dado que en (4.5) hay un número finito de sumandos. Mas precisamente, tenemos

$$\mathbf{E} X = \sum_{k=1}^m x_k p_k.$$

Si además se verifica $p_1 = \dots = p_m = 1/m$, entonces

$$\mathbf{E} X = \frac{1}{m} \sum_{k=1}^m x_k,$$

es decir, la esperanza matemática $\mathbf{E} X$ resulta ser el *promedio aritmético* de los valores que toma la variable aleatoria X . En general, en el caso discreto (4.5), se puede considerar a $\mathbf{E} X$ como un *promedio ponderado* de los valores x_k que toma la variable aleatoria X , donde las probabilidades p_k actúan como factores de ponderación.

Esperanza matemática para variables aleatorias con distribución absolutamente continua

Sea X una variable aleatoria absolutamente continua, con densidad $p(x)$. Sea $g(x)$ una función continua. Entonces

$$\mathbf{E} g(X) = \int_{-\infty}^{\infty} g(x)p(x)dx, \quad (4.6)$$

donde la esperanza existe si la integral anterior es absolutamente convergente, es decir, si $\int_{-\infty}^{\infty} |g(x)|p(x)dx < \infty$. En el caso particular en el que $g(x) = x$ para todo x real, obtenemos

$$\mathbf{E} X = \int_{-\infty}^{\infty} xp(x)dx, \quad (4.7)$$

si la integral anterior es absolutamente convergente. La fórmula (4.6) se obtiene a partir de (4.3).

Calculemos ahora las esperanzas matemáticas de algunas variables aleatorias cuyas distribuciones se encuentran en diversas aplicaciones.

Ejemplo 4.1. Calcular la esperanza matemática de una variable aleatoria X con distribución degenerada.

Tenemos $\mathbf{P}(X = c) = 1$, donde c es una constante dada. La variable aleatoria X tiene distribución discreta, y toma el único valor c con probabilidad 1. Aplicando (4.5), obtenemos

$$\mathbf{E} X = c \times 1 = c.$$

Ejemplo 4.2. Calcular la esperanza matemática de los puntos obtenidos al tirar un dado.

Tenemos que calcular $\mathbf{E} X$ para una variable aleatoria X con distribución discreta, que toma, con probabilidad $1/6$, cada uno de los valores 1, 2, 3, 4, 5 ó 6. Aplicando la fórmula (4.5), tenemos

$$\mathbf{E} X = \frac{1}{6} \sum_{k=1}^6 k = \frac{21}{6} = 3,5.$$

Ejemplo 4.3. Calcular la esperanza matemática de una variable aleatoria X con distribución binomial con parámetros (n, p) .

Esta variable aleatoria toma el valor k , con probabilidad $\mathbf{P}(X = k) = \binom{n}{k} p^k q^{n-k}$, donde $q = 1 - p$, para $k = 0, 1, \dots, n$. Aplicando (4.5), tenemos

$$\begin{aligned} \mathbf{E} X &= \sum_{k=0}^n k \mathbf{P}(X = k) = \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k} = np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i q^{n-1-i}, \end{aligned}$$

donde pusimos $i = k - 1$. Como la suma última vale $(q + p)^{n-1} = 1$, obtenemos que $\mathbf{E} X = np$.

Ejemplo 4.4. Calcular la esperanza matemática de una variable aleatoria X con distribución de Poisson con parámetro $\lambda > 0$.

La variable aleatoria X toma el valor k con probabilidad $\mathbf{P}(X = k) = \lambda^k e^{-\lambda} / k!$, para $k = 0, 1, \dots$. Aplicando la fórmula (4.5), tenemos

$$\mathbf{E} X = \sum_{k=0}^{\infty} k \mathbf{P}(X = k) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Ejemplo 4.5. Calcular la esperanza matemática de una variable aleatoria X con distribución normal con parámetros (a, σ) .

La variable aleatoria considerada tiene densidad, dada por

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

Aplicando la fórmula (4.7), tenemos

$$\mathbf{E} X = \int_{-\infty}^{\infty} xp(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-(x-a)^2/(2\sigma^2)} dx.$$

Al realizar el cambio de variable $u = (x - a)/\sigma$, obtenemos

$$\begin{aligned} \mathbf{E} X &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (a + \sigma u) e^{-u^2/2} du \\ &= \frac{a}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u e^{-u^2/2} du = a, \end{aligned}$$

en vista de que la primer integral vale $\sqrt{2\pi}$, y la segunda es nula, por ser el integrando una función impar.

Ejemplo 4.6. Calcular la esperanza matemática de una variable aleatoria X con distribución de Cauchy.

La variable aleatoria considerada tiene densidad, dada por

$$p(x) = 1/(\pi(1 + x^2)).$$

La esperanza matemática de esta variable existe, si $\int_{-\infty}^{\infty} |x|p(x)dx < \infty$. Tenemos

$$\int_0^b xp(x) = \frac{1}{\pi} \int_0^b \frac{x}{1+x^2} dx = \frac{1}{2\pi} \ln(1+x^2) \Big|_0^b = \frac{1}{2\pi} \ln(1+b^2) \rightarrow \infty,$$

si $b \rightarrow \infty$. En conclusión, $\int_{-\infty}^{\infty} |x|p(x)dx = 2 \int_0^{\infty} xp(x)dx = \infty$. De esta forma, la esperanza matemática de la variable aleatoria considerada no existe.

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y un vector aleatorio (X, Y) con función de distribución $F(x, y)$. Argumentos análogos a los del principio de esta sección permiten afirmar que el vector aleatorio $(X, Y): \Omega \rightarrow \mathbb{R}^2$ genera un nuevo espacio de probabilidad $(\mathbb{R}^2, \mathcal{B}, \mathbf{P}_{XY})$, donde \mathcal{B} es la σ -álgebra de los conjuntos de Borel en \mathbb{R}^2 , y \mathbf{P}_{XY} es una medida de probabilidad, definida a partir de la igualdad $\mathbf{P}_{XY}(D) = \mathbf{P}((X, Y) \in D)$ para cualquier rectángulo D del plano \mathbb{R}^2 .

Tiene lugar entonces la siguiente igualdad, análoga a (4.2):

$$\int_{\Omega} g(X, Y) d\mathbf{P} = \int_{\mathbb{R}^2} g(x, y) d\mathbf{P}_{XY}, \quad (4.8)$$

para una función continua $g(x, y)$ (y más en general, para funciones $g(x, y)$ de Borel).

Si el vector aleatorio tiene distribución discreta, y toma los valores (x_k, y_j) con probabilidades $p_{k,j} = \mathbf{P}(X = x_k, Y = y_j)$ ($k, j = 1, 2, \dots$), aplicando la fórmula (4.8) se obtiene, que

$$\mathbf{E} g(X, Y) = \sum_{k,j} g(x_k, y_j) p_{k,j}, \quad (4.9)$$

si la serie anterior es absolutamente convergente.

Si el vector aleatorio tiene densidad $p(x, y)$, entonces

$$\mathbf{E} g(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p(x, y) dx dy, \quad (4.10)$$

si la integral anterior es absolutamente convergente.

Estudiamos ahora las propiedades de la esperanza matemática.

Propiedad 1. Consideremos una variable aleatoria X con esperanza³ matemática $\mathbf{E}X$, y dos reales a, b . Entonces $\mathbf{E}(aX + b) = a\mathbf{E}X + b$.

Esta proposición es una consecuencia directa de la definición (4.1). Se puede también obtener, a partir de las fórmulas (4.5) y (4.7), correspondientes a las distribuciones discretas y absolutamente continuas.

Propiedad 2. Consideremos dos variables aleatorias X, Y con esperanzas respectivas $\mathbf{E}X, \mathbf{E}Y$. Entonces

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y.$$

Demostración. Esta proposición es consecuencia inmediata de la fórmula (4.1). Veamos la demostración cuando las variables tienen distribución discreta; X toma los valores x_1, x_2, \dots ; Y toma los valores y_1, y_2, \dots . Aplicando la fórmula (4.9), tenemos

$$\begin{aligned} \mathbf{E}(X + Y) &= \sum_{i,j} (x_i + y_j) \mathbf{P}(X = x_i, Y = y_j) \\ &= \sum_i x_i \sum_j \mathbf{P}(X = x_i, Y = y_j) \\ &\quad + \sum_j y_j \sum_i \mathbf{P}(X = x_i, Y = y_j) \\ &= \sum_i x_i \mathbf{P}(X = x_i) + \sum_j y_j \mathbf{P}(Y = y_j) = \mathbf{E}X + \mathbf{E}Y, \end{aligned}$$

lo que concluye la demostración en el caso discreto. Si el vector aleatorio (X, Y) tiene densidad $p(x, y)$, aplicando la fórmula (4.10) obtenemos

$$\begin{aligned} \mathbf{E}(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)p(x, y)dx dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} p(x, y)dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} p(x, y)dx \right) dy \\ &= \int_{-\infty}^{\infty} xp_1(x)dx + \int_{-\infty}^{\infty} yp_2(y)dy = \mathbf{E}X + \mathbf{E}Y, \end{aligned}$$

donde $p_1(x)$ y $p_2(x)$ son las densidades de las variables aleatorias X e Y respectivamente. \square

³Se entiende que existe la esperanza matemática de la variable aleatoria.

Propiedad 3. Consideremos dos variables aleatorias independientes X, Y con esperanzas respectivas $\mathbf{E}X, \mathbf{E}Y$. Entonces

$$\mathbf{E}(XY) = \mathbf{E}X \mathbf{E}Y. \quad (4.11)$$

Demostración. Demostremos esta proposición en primer lugar, en el caso en el que las variables aleatorias tienen distribución discreta; X toma los valores x_1, x_2, \dots ; Y toma los valores y_1, y_2, \dots . Aplicando (4.9) y la proposición 3.1, tenemos

$$\begin{aligned} \mathbf{E}(XY) &= \sum_{i,j} x_i y_j \mathbf{P}(X = x_i, Y = y_j) = \sum_{i,j} x_i y_j \mathbf{P}(X = x_i) \mathbf{P}(Y = y_j) \\ &= \sum_i x_i \mathbf{P}(X = x_i) \sum_j y_j \mathbf{P}(Y = y_j) = \mathbf{E}X \mathbf{E}Y. \end{aligned}$$

Si el vector aleatorio (X, Y) tiene densidad $p(x, y)$, aplicando la proposición 3.3, obtenemos que $p(x, y) = p_1(x)p_2(y)$, donde $p_1(x)$ y $p_2(y)$ son las densidades de las variables aleatorias X e Y respectivamente. En vista de (4.10), tenemos

$$\begin{aligned} \mathbf{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp_1(x)p_2(y) dx dy \\ &= \int_{-\infty}^{\infty} xp_1(x) dx \int_{-\infty}^{\infty} yp_2(y) dy = \mathbf{E}X \mathbf{E}Y, \end{aligned}$$

concluyendo la demostración en el caso absolutamente continuo.

Veamos ahora una demostración en el caso general. Sabemos que el resultado es cierto en el caso discreto. Supongamos primero que X e Y son variables aleatorias independientes y no negativas, con esperanzas respectivas $\mathbf{E}X$ y $\mathbf{E}Y$, definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$.

Consideremos para cada $n = 1, 2, \dots$, las variables aleatorias de la forma

$$X_n(\omega) = \begin{cases} (i-1)/2^n, & \text{si } (i-1)/2^n \leq X(\omega) < i/2^n \quad (i = 1, \dots, n2^n), \\ 0, & \text{si } X(\omega) \geq n. \end{cases}$$

Es claro que $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ para todo $\omega \in \Omega$, donde la sucesión $\{X_n(\omega)\}$ es no decreciente. En forma análoga introducimos la sucesión $\{Y_n(\omega)\}$, que presenta el mismo comportamiento. Las variables aleatorias X_n e Y_n tienen distribución discreta, y toman una cantidad finita de

valores. Por ser X e Y independientes, aplicando la proposición 3.1, las variables X_n e Y_n resultan ser, también, independientes. Entonces, por lo ya demostrado

$$\mathbf{E}(X_n Y_n) = \mathbf{E} X_n \mathbf{E} Y_n. \quad (4.12)$$

Teniendo en cuenta, que $X_n(\omega) \rightarrow X(\omega)$ para todo $\omega \in \Omega$, obtenemos⁴ $\int_{\Omega} X_n d\mathbf{P} \rightarrow \int_{\Omega} X d\mathbf{P}$, es decir, $\mathbf{E} X_n \rightarrow \mathbf{E} X$. En forma análoga obtenemos $\mathbf{E} Y_n \rightarrow \mathbf{E} Y$, y también $\mathbf{E}(X_n Y_n) \rightarrow \mathbf{E}(XY)$. Entonces, por la fórmula (4.12), concluimos que $\mathbf{E}(XY) = \mathbf{E} X \mathbf{E} Y$.

Sean ahora X e Y variables aleatorias arbitrarias, para las cuales existe la esperanza matemática. Definimos

$$X^+(\omega) = \text{máx}(0, X(\omega)), \quad X^-(\omega) = \text{máx}(0, -X(\omega)).$$

Es claro que $X^+(\omega) \geq 0$, $X^-(\omega) \geq 0$, y que $X(\omega) = X^+(\omega) - X^-(\omega)$ para cada $\omega \in \Omega$. Definimos en forma análoga las variables aleatorias Y^+ e Y^- .

Veamos que las variables aleatorias X^+ e Y^+ son independientes. Supongamos primero que $x \geq 0$, e $y \geq 0$. En este caso, tenemos las igualdades de sucesos $\{X^+ \leq x\} = \{X \leq x\}$, $\{Y^+ \leq y\} = \{Y \leq y\}$, y por ésto,

$$\begin{aligned} \mathbf{P}(X^+ \leq x, Y^+ \leq y) &= \mathbf{P}(X \leq x, Y \leq y) \\ &= \mathbf{P}(X \leq x) \mathbf{P}(Y \leq y) = \mathbf{P}(X^+ \leq x) \mathbf{P}(Y^+ \leq y), \end{aligned}$$

donde utilizamos que las variables aleatorias X e Y son independientes. Veamos las posibilidades restantes. Si $x < 0$, el suceso $\{X^+ \leq x\}$ es imposible, por lo que $\mathbf{P}(X^+ \leq x, Y^+ \leq y) = \mathbf{P}(\emptyset) = 0 = \mathbf{P}(X^+ \leq x) \mathbf{P}(Y^+ \leq y)$. Si $y < 0$, ocurre lo mismo. En conclusión, las variables aleatorias X^+ e Y^+ son independientes. Es análogo demostrar la independencia de las parejas de variables aleatorias X^+ e Y^- , X^- e Y^+ , X^- e Y^- . Por ésto

$$\begin{aligned} \mathbf{E}(XY) &= \mathbf{E}(X^+ - X^-)(Y^+ - Y^-) \\ &= \mathbf{E}(X^+ Y^+ - X^+ Y^- - X^- Y^+ + X^- Y^-) \\ &= \mathbf{E}(X^+ Y^+) - \mathbf{E}(X^+ Y^-) - \mathbf{E}(X^- Y^+) + \mathbf{E}(X^- Y^-) \\ &= \mathbf{E} X^+ \mathbf{E} Y^+ - \mathbf{E} X^+ \mathbf{E} Y^- - \mathbf{E} X^- \mathbf{E} Y^+ + \mathbf{E} X^- \mathbf{E} Y^- \\ &= (\mathbf{E} X^+ - \mathbf{E} X^-)(\mathbf{E} Y^+ - \mathbf{E} Y^-) = \mathbf{E} X \mathbf{E} Y. \end{aligned}$$

concluyendo la demostración. □

⁴Aquí se utiliza el teorema de convergencia monótona.

4.2. Varianza

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y una variable aleatoria $X = X(\omega)$, $\omega \in \Omega$. Supongamos que existe la esperanza matemática de esta variable aleatoria, dada por $\mathbf{E} X = a$. Definimos la *varianza* (llamada también *variancia*) de la variable aleatoria X , como la esperanza matemática de la variable aleatoria $(X - a)^2$, cuando esta última esperanza existe, y la designamos $\mathbf{var} X$. De esta forma, hemos definido

$$\mathbf{var} X = \mathbf{E}(X - a)^2, \quad (4.13)$$

si la esperanza anterior existe, donde suponemos que existe la esperanza matemática de la variable aleatoria X , dada por

$$a = \mathbf{E} X.$$

Al valor positivo de la raíz cuadrada de la varianza le llamamos *desviación estándar* de la variable aleatoria X , y lo designamos σ_X . De esta forma, $\sigma_X = \sqrt{\mathbf{var} X}$, o en forma equivalente, $\mathbf{var} X = \sigma_X^2$.

Según nuestra definición de esperanza matemática (4.1) y la identidad (4.2), en el caso general se obtienen las fórmulas

$$\mathbf{var} X = \int_{\Omega} (X - a)^2 d\mathbf{P} = \int_{-\infty}^{\infty} (x - a)^2 dF(x), \quad (4.14)$$

y la varianza existe si alguna de las integrales anteriores es convergente (lo que equivale, en este caso, a la convergencia absoluta), donde suponemos que existe la esperanza

$$a = \mathbf{E} X = \int_{\Omega} X d\mathbf{P} = \int_{-\infty}^{\infty} x dF(x).$$

Consideremos ahora los dos tipos más importantes de distribuciones. Si la variable aleatoria X tiene distribución discreta, y toma los valores x_1, x_2, \dots con probabilidades p_1, p_2, \dots respectivamente, entonces

$$\mathbf{var} X = \sum_k (x_k - a)^2 p_k, \quad (4.15)$$

si la serie anterior es convergente, donde suponemos que existe la esperanza

$$a = \sum_k x_k p_k.$$

Si X es una variable aleatoria con densidad $p(x)$, entonces

$$\mathbf{var} X = \int_{-\infty}^{\infty} (x - a)^2 p(x) dx, \quad (4.16)$$

si la integral anterior es convergente, donde suponemos que existe la esperanza

$$a = \int_{-\infty}^{\infty} xp(x) dx.$$

Las fórmulas (4.15) y (4.16) se deducen de la definición de varianza (4.13), y de las fórmulas (4.4) y (4.6), donde ponemos $g(x) = (x - a)^2$.

Veamos dos identidades útiles, que verifica la varianza de una variable aleatoria arbitraria X . Tenemos

$$\mathbf{var} X = \mathbf{E}(X^2) - (\mathbf{E} X)^2, \quad (4.17)$$

$$\mathbf{var} X = \mathbf{E}(X(X - 1)) - \mathbf{E} X(\mathbf{E} X - 1). \quad (4.18)$$

Demostremos (4.17). Aplicando la definición de varianza (4.13), obtenemos

$$\mathbf{var} X = \mathbf{E}(X^2 - 2aX + a^2) = \mathbf{E}(X^2) - 2a \mathbf{E} X + a^2 = \mathbf{E}(X^2) - (\mathbf{E} X)^2.$$

Para demostrar (4.18), tenemos

$$\begin{aligned} \mathbf{E}(X(X - 1)) - \mathbf{E} X(\mathbf{E} X - 1) &= \mathbf{E}(X^2) - \mathbf{E} X - (\mathbf{E} X)^2 + \mathbf{E} X \\ &= \mathbf{E}(X^2) - (\mathbf{E} X)^2 = \mathbf{var} X. \end{aligned}$$

Ejemplo 4.7. Calcular la varianza de una variable aleatoria X con distribución binomial con parámetros (n, p) .

Sabemos del ejemplo 4.3 que $\mathbf{E} X = np$. Utilicemos la identidad (4.18), especialmente adecuada si X toma únicamente valores enteros. Si $q = 1 - p$, tenemos

$$\begin{aligned} \mathbf{E}(X(X - 1)) &= \sum_{k=0}^n k(k - 1) \mathbf{P}(X = k) = \sum_{k=0}^n k(k - 1) \binom{n}{k} p^k q^{n-k} \\ &= n(n - 1)p^2 \sum_{i=0}^{n-2} \frac{(n - 2)!}{i!(n - 2 - i)!} p^i q^{n-2-i} \\ &= n(n - 1)p^2 (q + p)^{n-2} = n(n - 1)p^2. \end{aligned}$$

Por ésto, de (4.18) obtenemos

$$\mathbf{var} X = n(n - 1)p^2 - np(np - 1) = np(1 - p) = npq.$$

Ejemplo 4.8. Calcular la varianza de una variable aleatoria X con distribución de Poisson con parámetro $\lambda > 0$.

Tenemos

$$\begin{aligned}\mathbf{E}(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2.\end{aligned}$$

Por ésto, teniendo en cuenta que $\mathbf{E}X = \lambda$ (ver ejemplo 4.4) y la fórmula (4.18), obtenemos

$$\mathbf{var} X = \lambda^2 - \lambda(\lambda - 1) = \lambda.$$

Ejemplo 4.9. Calcular la varianza de una variable aleatoria X con distribución normal con parámetros (a, σ) .

En el ejemplo 4.5 obtuvimos que $\mathbf{E}X = a$. Aplicando (4.16), tenemos

$$\mathbf{var} X = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/(2\sigma^2)} dx.$$

Introducimos el cambio de variable $u = (x-a)/\sigma$ y aplicamos la fórmula de integración por partes, para obtener

$$\begin{aligned}\mathbf{var} X &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2/2} du = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ud(-e^{-u^2/2}) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = \sigma^2.\end{aligned}$$

De esta forma $\mathbf{var} X = \sigma^2$, y en consecuencia σ es la desviación estándar de la variable aleatoria X . Encontramos así el sentido probabilístico de los parámetros a y σ que caracterizan a una distribución normal.

Veamos las propiedades que verifica la varianza de una variable aleatoria.

Propiedad 1. *Para cualquier variable aleatoria X se verifica $\mathbf{var} X \geq 0$.*

Esta conclusión es inmediata a partir de cualquiera de las definiciones que se utilice.

Propiedad 2. *Se verifica $\mathbf{var} X = 0$ si y solo si la variable aleatoria X tiene distribución degenerada.*

Demostración. Supongamos que X tiene distribución degenerada, es decir $\mathbf{P}(X = c) = 1$ para alguna constante c real. Entonces $\mathbf{E} X = c$, y $\mathbf{var} X = \mathbf{E}(X - c)^2 = (c - c)^2 \times 1 = 0$.

Supongamos ahora que $\mathbf{var} X = 0$ y sea $a = \mathbf{E} X$. Demostremos primero que X tiene distribución degenerada cuando tiene distribución discreta, y toma los valores x_1, x_2, \dots con probabilidades p_1, p_2, \dots respectivamente. En este caso tenemos

$$\mathbf{var} X = \sum_k (x_k - a)^2 p_k = 0.$$

Entonces cada sumando es nulo, y $x_k = a$ para todo $k = 1, 2, \dots$ (porque $p_k > 0$). Por ésto $\mathbf{P}(X = a) = p_1 + p_2 + \dots = 1$, y X tiene distribución degenerada.

Veamos la demostración en el caso general, si el lector considera posible utilizar la definición de varianza (4.14). En este caso tenemos

$$\mathbf{var} X = \int_{\Omega} (X - a)^2 d\mathbf{P} = 0.$$

Como $(X - a)^2 \geq 0$, entonces $\mathbf{P}((X(\omega) - a)^2 = 0) = 1$ es decir, $\mathbf{P}(X = a) = 1$, y la variable aleatoria X tiene distribución degenerada. \square

Propiedad 3. *Sea X una variable aleatoria con varianza $\mathbf{var} X$, y sean a, b constantes. Entonces*

$$\mathbf{var}(aX + b) = a^2 \mathbf{var} X.$$

Demostración. Por la definición de varianza, tenemos

$$\begin{aligned} \mathbf{var}(aX + b) &= \mathbf{E}(aX + b - \mathbf{E}(aX + b))^2 = \mathbf{E}(aX + b - a\mathbf{E} X - b)^2 \\ &= \mathbf{E}(a(X - \mathbf{E} X))^2 = a^2 \mathbf{E}(X - \mathbf{E} X)^2 = a^2 \mathbf{var} X, \end{aligned}$$

concluyendo la demostración. \square

Propiedad 4. *Sean X e Y variables aleatorias independientes, para las cuales existe la varianza. Entonces*

$$\mathbf{var}(X + Y) = \mathbf{var} X + \mathbf{var} Y.$$

Esta propiedad es un caso particular de la siguiente.

Propiedad 5. Sean X_1, \dots, X_n variables aleatorias independientes dos a dos, para las cuales existe la varianza. Entonces

$$\mathbf{var}(X_1 + \dots + X_n) = \mathbf{var} X_1 + \dots + \mathbf{var} X_n. \quad (4.19)$$

Demostración. Designemos $a_k = \mathbf{E} X_k$ ($k = 1, \dots, n$). Partiendo de la definición de varianza, tenemos

$$\begin{aligned} \mathbf{var} \sum_{k=1}^n X_k &= \mathbf{E} \left(\sum_{k=1}^n X_k - \mathbf{E} \sum_{k=1}^n X_k \right)^2 = \mathbf{E} \left(\sum_{k=1}^n (X_k - a_k) \right)^2 \\ &= \sum_{k=1}^n \mathbf{E} (X_k - a_k)^2 + 2 \sum_{1 \leq k < j \leq n} \mathbf{E} (X_k - a_k)(X_j - a_j) \\ &= \sum_{k=1}^n \mathbf{var} X_k, \end{aligned}$$

porque

$$\mathbf{E} (X_k - a_k)(X_j - a_j) = \mathbf{E} (X_k - a_k) \mathbf{E} (X_j - a_j) = 0.$$

en vista de la independencia de las variables aleatorias X_k y X_j para $j \neq k$. Esto concluye la demostración. \square

Combinando las propiedades 3 y 5 se obtiene la siguiente proposición: si X_1, \dots, X_n son variables independientes dos a dos, para las cuales existe la varianza, y a_1, \dots, a_n, b son constantes, entonces

$$\mathbf{var}(a_1 X_1 + \dots + a_n X_n + b) = a_1^2 \mathbf{var} X_1 + \dots + a_n^2 \mathbf{var} X_n.$$

En particular, si X e Y son variables aleatorias independientes para las cuales existe la varianza, tenemos $\mathbf{var}(X - Y) = \mathbf{var} X + \mathbf{var} Y$.

La esperanza matemática y la varianza son las dos características numéricas más importantes de una variable aleatoria; otras características numéricas serán consideradas en la sección 4. La esperanza matemática de una variable aleatoria caracteriza la posición central, o representativa, de los valores que ésta toma. Por su parte, la varianza es una característica del grado de concentración de estos valores alrededor de la esperanza. Cuanto mayor sea esta concentración, menor será la varianza (en otras palabras: cuanto mayor sea la dispersión de los valores que toma la variable aleatoria alrededor de la esperanza, mayor será la varianza).

La esperanza matemática de una variable aleatoria es una característica más precisa, cuanto menor es su varianza.

Sea X es una variable aleatoria con función de distribución $F(x)$. Supongamos que en la recta real se distribuye una unidad de masa, de acuerdo a la siguiente regla: a la izquierda de cada punto x se encuentra una cantidad de masa igual a $F(x)$. Entonces, $\mathbf{E} X$ indica la abscisa del centro de gravedad de esta masa distribuida, y $\mathbf{var} X$ su momento de inercia.

4.3. Desigualdad de Chebishev

Comencemos demostrando la siguiente proposición.

Lema 4.1. *Consideremos una variable aleatoria X no negativa, para la cual existe la esperanza matemática $\mathbf{E} X$. Para cualquier $t > 0$, tenemos*

$$\mathbf{P}(X \geq t) \leq \frac{1}{t} \mathbf{E} X. \quad (4.20)$$

Demostración. Demostremos primero este lema cuando X tiene distribución discreta, y toma los valores x_1, x_2, \dots , con probabilidades p_1, p_2, \dots , respectivamente. Como $x_k \geq 0$ para todo k , tenemos

$$\mathbf{E} X = \sum_k x_k p_k \geq \sum_{k: x_k \geq t} x_k p_k \geq \sum_{k: x_k \geq t} t p_k = t \mathbf{P}(X \geq t),$$

y de aquí se obtiene (4.20).

Si X tiene densidad $p(x)$ resulta que $p(x) = 0$ para todo $x < 0$, porque $x < 0$ implica $F(x) = \mathbf{P}(X \leq x) = 0$ cuando $X \geq 0$. Por ésto,

$$\begin{aligned} \mathbf{E} X &= \int_{-\infty}^{\infty} x p(x) dx = \int_0^{\infty} x p(x) dx \geq \int_t^{\infty} x p(x) dx \geq \int_t^{\infty} t p(x) dx \\ &= t \int_t^{\infty} p(x) dx = t \mathbf{P}(X \geq t), \end{aligned}$$

obteniendo (4.20). Si el lector considera posible utilizar la definición (4.1), la demostración del lema sigue el mismo esquema:

$$\mathbf{E} X = \int_{\Omega} X d\mathbf{P} \geq \int_{\{\omega: X(\omega) \geq t\}} X d\mathbf{P} \geq t \int_{\{\omega: X(\omega) \geq t\}} d\mathbf{P} = t \mathbf{P}(X \geq t),$$

concluyendo la demostración. \square

Este lema fue demostrado por Chebishev. La condición $X \geq 0$ no se puede omitir; sin ella podría ocurrir que $\mathbf{E} X < 0$, lo que es incompatible con la desigualdad (4.20).

Sea ahora X una variable aleatoria arbitraria, para la cual existe la varianza. Designemos $Z = (X - \mathbf{E} X)^2$. Es claro que Z toma valores no negativos, y que $\mathbf{E} Z = \mathbf{var} X$. Por el lema anterior tenemos que para todo $\varepsilon > 0$, se verifica

$$\mathbf{P}((X - \mathbf{E} X)^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} \mathbf{var} X,$$

que, tomando raíz cuadrada, es

$$\mathbf{P}(|X - \mathbf{E} X| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{var} X. \quad (4.21)$$

La desigualdad (4.21), válida para cualquier variable con varianza y para cualquier $\varepsilon > 0$, se denomina *desigualdad de Chebishev*.

Esta desigualdad resulta ser una de las más utilizadas en la teoría de la probabilidad. Teniendo en cuenta, que la suma de las probabilidades de un suceso y su contrario es igual a uno, la desigualdad de Chebishev (4.21) se puede escribir de la forma

$$\mathbf{P}(|X - \mathbf{E} X| < \varepsilon) \geq 1 - \frac{1}{\varepsilon^2} \mathbf{var} X,$$

para cualquier $\varepsilon > 0$.

4.4. Momentos de órdenes superiores. Mediana y cuantiles

Consideremos una variable aleatoria X y un natural $k \geq 1$. Definimos el *momento de orden k* de la variable aleatoria X , también llamado *momento inicial de orden k* , que designamos α_k , mediante la identidad

$$\alpha_k = \mathbf{E}(X^k),$$

cuando la esperanza anterior existe. En particular, $\alpha_1 = \mathbf{E} X$. Observemos que la existencia del momento de orden k de una variable aleatoria es equivalente a la finitud del así llamado *momento absoluto de orden k* ,

también llamado *momento absoluto inicial de orden k* , designado β_k y definido mediante la identidad

$$\beta_k = \mathbf{E} |X|^k.$$

Si una variable aleatoria X está acotada, es decir, si $|X| \leq C$ para alguna constante C , entonces, existen sus momentos α_k y β_k para cualquier k . (Ver la observación luego de la definición (4.1), página 87.)

Proposición 4.1. *Sea X una variable aleatoria. Si existe su momento de orden k , entonces, existe su momento de orden m , para todo $m = 1, \dots, k$.*

Demostración. Elegimos $1 \leq m \leq k$. Vale la desigualdad $|X|^m \leq 1 + |X|^k$. En efecto, si $|X| \leq 1$, se tiene $|X|^m \leq 1$, y si $|X| \geq 1$, se tiene $|X|^m \leq |X|^k$. De esta desigualdad se obtiene que $\mathbf{E} |X|^m \leq 1 + \mathbf{E} |X|^k$, es decir, $\beta_m \leq 1 + \beta_k$ si $1 \leq m \leq k$. Como β_k es finito, β_m es finito, concluyendo la demostración. \square

Consideremos una variable aleatoria X y un natural $k \geq 1$. Definimos el *momento centrado de orden k* de la variable aleatoria X , que designamos μ_k , mediante la identidad

$$\mu_k = \mathbf{E}(X - \mathbf{E} X)^k,$$

cuando la esperanzas anteriores existen.

Como ocurre con el momento de orden k , la existencia del momento centrado de orden k de una variable aleatoria es equivalente a la finitud del *momento absoluto centrado de orden k* , designado ν_k y definido mediante la identidad

$$\nu_k = \mathbf{E} |X - \mathbf{E} X|^k.$$

cuando la esperanza anterior es finita. En particular, $\mu_2 = \nu_2 = \mathbf{var} X$.

Consideremos un número real $p \geq 1$. Definimos el *momento absoluto inicial* y el *momento absoluto centrado* de orden p de una variable aleatoria X , mediante las fórmulas

$$\beta_p = \mathbf{E} |X|^p, \quad \nu_p = \mathbf{E} |X - \mathbf{E} X|^p,$$

respectivamente, si existen las esperanzas. De la desigualdad de Cauchy-Bunyakovsky se obtiene, que $\mathbf{E} |X| \leq \sqrt{\mathbf{E}(X^2)}$, es decir $\beta_1 \leq \beta_2^{1/2}$, si los momentos anteriores existen. Esta desigualdad se obtiene también de la

desigualdad de Lyapunov: $\beta_s^{1/s} \leq \beta_r^{1/r}$, si $0 < s < r$. (Estas desigualdades se presentan sin aquí demostración, ver ejercicio 7, capítulo 9.)

Definimos la *función generatriz de momentos* de una variable aleatoria X , mediante la identidad

$$M(t) = \mathbf{E} e^{tX},$$

para los valores reales de t para los cuales existe la esperanza anterior. Si bien esta esperanza siempre existe para $t = 0$, no siempre se tiene un intervalo no degenerado que contenga al origen, en donde $M(t)$ esté definida. Si esta función está definida para todo t de un intervalo $|t| \leq b$, en este intervalo vale la fórmula

$$M(t) = 1 + \sum_{k=1}^{\infty} \alpha_k \frac{t^k}{k!},$$

que surge de su desarrollo de Taylor. Si $|X| \leq C$ para alguna constante C , entonces, la función generatriz de momentos está definida para todo t real.

Como los momentos de una variable aleatoria no siempre existen, se introducen otras características numéricas de las variables aleatorias, que existen siempre.

Consideremos una variable aleatoria X y un número $0 < q < 1$. Llamamos *cuantil* de orden q de la variable aleatoria X , a cualquier número κ_q que verifique las condiciones

$$\mathbf{P}(X \leq \kappa_q) \geq q, \quad \mathbf{P}(X \geq \kappa_q) \geq 1 - q.$$

Si la función de distribución de una variable aleatoria X es estrictamente creciente en toda la recta real, su cuantil de cualquier orden q es único. Una variable aleatoria arbitraria no siempre cumple esta condición.

El cuantil de orden $1/2$ se denomina *mediana*. De esta forma, la mediana de una variable aleatoria X es cualquier número m , que verifica las condiciones

$$\mathbf{P}(X \leq m) \geq \frac{1}{2}, \quad \mathbf{P}(X \geq m) \geq \frac{1}{2}.$$

Ejemplo 4.10. Consideremos una variable aleatoria X , que toma el valor 0 con probabilidad $1/2$, y el valor 1 con probabilidad $1/2$. Cualquier número del intervalo cerrado $[0, 1]$ es mediana de esta variable aleatoria.

Los cuantiles de orden $1/4$, $1/2$ y $3/4$, se denominan *cuartiles*.

La *moda* de una variable aleatoria X se define únicamente para variables aleatorias con distribución discreta o absolutamente continua. Si X tiene distribución discreta, y toma los valores $x_1 < x_2 < x_3 < \dots$, con probabilidades p_1, p_2, p_3, \dots , respectivamente, entonces, la *moda* de la variable aleatoria X es cualquier número x_k para el que se cumpla $p_{k-1} \leq p_k \geq p_{k+1}$. Si X tiene densidad dada por $p(x)$, entonces, la *moda* de X es cualquier punto de máximo local de $p(x)$. Es claro, que en general, la moda no es única.

Si una variable aleatoria tiene una única moda su distribución se denomina *unimodal*.

Para una variable aleatoria X con distribución normal con parámetros (a, σ) , la esperanza matemática, la mediana y la moda coinciden, y toman el valor del parámetro a . La mediana y la moda en este caso son únicas.

4.5. Covarianza, coeficiente de correlación. Matriz de Covarianza

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, y dos variables aleatorias X e Y , con esperanzas matemáticas $\mathbf{E}X$ y $\mathbf{E}Y$ respectivamente. Definimos la *covarianza* entre X e Y como la esperanza matemática del producto $(X - \mathbf{E}X)(Y - \mathbf{E}Y)$, cuando esta esperanza existe, y la designamos $\mathbf{cov}(X, Y)$. Hemos entonces definido

$$\mathbf{cov}(X, Y) = \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y), \quad (4.22)$$

cuando la esperanza del producto $(X - \mathbf{E}X)(Y - \mathbf{E}Y)$ existe. Como

$$\begin{aligned} \mathbf{E}(X - \mathbf{E}X)(Y - \mathbf{E}Y) &= \mathbf{E}(XY - X\mathbf{E}Y - Y\mathbf{E}X + \mathbf{E}X\mathbf{E}Y) \\ &= \mathbf{E}(XY) - \mathbf{E}Y\mathbf{E}X - \mathbf{E}X\mathbf{E}Y + \mathbf{E}X\mathbf{E}Y \\ &= \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y, \end{aligned}$$

tiene lugar la identidad

$$\mathbf{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y. \quad (4.23)$$

Si las variables aleatorias son independientes, aplicando la propiedad 3 se obtiene que $\mathbf{cov}(X, Y) = 0$. Además, $\mathbf{cov}(X, X) = \mathbf{var} X$, en vista de las definiciones de varianza y covarianza.

Proposición 4.2. Si X_1, \dots, X_n son variables aleatorias, entonces

$$\mathbf{var} \sum_{k=1}^n X_k = \sum_{k=1}^n \mathbf{var} X_k + 2 \sum_{1 \leq k < j \leq n} \mathbf{cov}(X_k, X_j) \quad (4.24)$$

siempre que las varianzas y covarianzas a la derecha de la fórmula anterior existan⁵.

Para demostrar esta proposición es suficiente observar que la fórmula (4.24) fue obtenida en el curso de la demostración de la propiedad 5. Si suponemos además que las variables X_1, \dots, X_n son independientes dos a dos, entonces $\mathbf{cov}(X_k, X_j) = 0$ para $k \neq j$, y obtenemos la fórmula

$$\mathbf{var} \sum_{k=1}^n X_k = \sum_{k=1}^n \mathbf{var} X_k,$$

es decir, la propiedad 5. De esta forma, la propiedad 5 es un corolario de la proposición 4.2.

Consideremos dos variables aleatorias X e Y , con varianzas positivas, que designamos σ_X^2 y σ_Y^2 respectivamente. El *coeficiente de correlación* entre las variables aleatorias X e Y , designado $\rho(X, Y)$, se define mediante la identidad

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Según vimos, si X e Y son variables aleatorias independientes, entonces $\rho(X, Y) = 0$, porque $\mathbf{cov}(X, Y) = 0$. Si se verifica $\rho(X, Y) = 0$, decimos que las variables aleatorias X e Y son *no correlacionadas*. De esta forma, si dos variables aleatorias tienen varianza y son independientes, resultan ser no correlacionadas. El recíproco de esta afirmación en general no es cierto, como se ve a continuación.

Ejemplo 4.11. Consideremos una variable aleatoria X , que toma, con probabilidad $1/4$, cada uno de los valores $-2, -1, 1$ y 2 . Sea $Y = X^2$. Es claro que la variable aleatoria Y toma, con probabilidad $1/2$, cada uno de los valores 1 y 4 . La variable aleatoria $XY = X^3$ toma, con probabilidad $1/4$, cada uno de los valores $-8, -1, 1$ y 8 , por lo que $\mathbf{E}(XY) = 0$. Como $\mathbf{E}X = 0$, de la fórmula (4.23), obtenemos $\mathbf{cov}(X, Y) = 0$, y por lo tanto,

⁵En vista de la desigualdad de Cauchy-Bunyakovsky, es suficiente suponer que existen las varianzas de las variables aleatorias X_k ($k = 1, \dots, n$).

$\rho(X, Y) = 0$. Las variables X e Y son entonces no correlacionadas, pero no resultan independientes, dado que una de ellas es función de la otra: $Y = X^2$.

Proposición 4.3. *Consideremos dos variables aleatorias X e Y , para las cuales existe la varianza. Entonces, vale la desigualdad*

$$-1 \leq \rho(X, Y) \leq 1.$$

Demostración. Introducimos la variable aleatoria

$$Z = \frac{X - \mathbf{E} X}{\sigma_X} + \frac{Y - \mathbf{E} Y}{\sigma_Y}, \quad (4.25)$$

donde σ_X y σ_Y designan las desviaciones estándar de X e Y respectivamente. Como la esperanza matemática de una variable aleatoria no negativa es no negativa, tenemos $\mathbf{E}(Z^2) \geq 0$. Entonces, tenemos

$$\begin{aligned} 0 \leq \mathbf{E}(Z^2) &= \mathbf{E}\left(\frac{X - \mathbf{E} X}{\sigma_X} + \frac{Y - \mathbf{E} Y}{\sigma_Y}\right)^2 \\ &= \mathbf{E}\left(\frac{(X - \mathbf{E} X)^2}{\sigma_X^2} + \frac{(Y - \mathbf{E} Y)^2}{\sigma_Y^2} + 2\frac{(X - \mathbf{E} X)(Y - \mathbf{E} Y)}{\sigma_X \sigma_Y}\right) \\ &= \frac{1}{\sigma_X^2} \mathbf{var} X + \frac{1}{\sigma_Y^2} \mathbf{var} Y + 2\rho(X, Y) = 2(1 + \rho(X, Y)), \end{aligned} \quad (4.26)$$

donde aplicamos las definiciones de varianza y de correlación. En conclusión $2 + 2\rho(X, Y) \geq 0$, por lo que $\rho(X, Y) \geq -1$.

Si en lugar del signo $+$ en la definición (4.25) de Z , ponemos el signo $-$, un argumento análogo, nos permiten obtener la desigualdad $2(1 - \rho(X, Y)) \geq 0$, o sea $\rho(X, Y) \leq 1$, concluyendo la demostración. \square

Proposición 4.4. *Consideremos dos variables aleatorias X e Y , para las cuales existe la varianza. Entonces, la igualdad $|\rho(X, Y)| = 1$ tiene lugar si y solo si existen dos constantes $a \neq 0$ y b , tales que*

$$\mathbf{P}(Y = aX + b) = 1. \quad (4.27)$$

Demostración. Supongamos primero que $|\rho(X, Y)| = 1$, y que por ejemplo vale $\rho(X, Y) = -1$. Consideremos la variable aleatoria Z definida en (4.25). Como obtuvimos en (4.26), vale $\mathbf{E}(Z^2) = 0$. Además, la variable aleatoria Z cumple $\mathbf{E} Z = 0$, por lo que $\mathbf{var} Z = \mathbf{E}(Z^2) = 0$. En vista de la

proposición 2 obtenemos que Z es una variable aleatoria con distribución degenerada, que cumple $\mathbf{P}(Z = 0) = 1$, porque $\mathbf{E} Z = 0$ (ver ejemplo 4.1). Como $Z = (X - \mathbf{E} X)/\sigma_X + (Y - \mathbf{E} Y)/\sigma_Y$, sustituyendo, obtenemos que $\mathbf{P}(Y = aX + b) = 1$, donde $a = -\sigma_Y/\sigma_X$ y $b = \mathbf{E} Y + \sigma_Y \mathbf{E} X/\sigma_X$. Si $\rho(X, Y) = 1$, un argumento análogo (cambiando el signo de $+$ por el signo de $-$ en la definición de Z), nos permite obtener la fórmula (4.27), donde $a = \sigma_Y/\sigma_X$, y $b = \mathbf{E} Y - \mathbf{E} X \sigma_Y/\sigma_X$.

Supongamos ahora que tiene lugar (4.27), donde $a \neq 0$. Tenemos

$$\begin{aligned} \mathbf{cov}(X, Y) &= \mathbf{cov}(X, aX + b) = \mathbf{E} X(aX + b) - \mathbf{E} X \mathbf{E}(aX + b) \\ &= a \mathbf{E}(X^2) + b \mathbf{E} X - a(\mathbf{E} X)^2 - b \mathbf{E} X = a \mathbf{var} X = a\sigma_X^2. \end{aligned}$$

Por otra parte, $\sigma_Y^2 = \mathbf{var} Y = \mathbf{var}(aX + b) = a^2 \mathbf{var} X = a^2 \sigma_X^2$, según la propiedad 3. En conclusión, tenemos

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{a\sigma_X^2}{|a|\sigma_X^2} = \frac{a}{|a|} = \begin{cases} 1, & \text{si } a > 0, \\ -1, & \text{si } a < 0. \end{cases}$$

concluyendo la demostración. □

Ejemplo 4.12. Calcular el coeficiente de correlación $\rho(X, Y)$ entre dos variables aleatorias X e Y con distribución normal bidimensional, y densidad $p(x, y)$ dada en la fórmula (3.22).

Sabemos que X tiene distribución normal con parámetros (a_1, σ_1) ; Y tiene distribución normal con parámetros (a_2, σ_2) (ejemplo 3.15). Por ésto, tenemos $\mathbf{E} X = a_1$, $\mathbf{E} Y = a_2$, $\mathbf{var} X = \sigma_1^2$, $\mathbf{var} Y = \sigma_2^2$. Aplicando la definición de covarianza (4.22), tenemos

$$\begin{aligned} \mathbf{cov}(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - a_1)(y - a_2)p(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} (x - a_1)p(x, y) dx \right] (y - a_2) dy. \end{aligned}$$

Los cálculos que restan son similares a los del ejemplo 3.15. Para calcular la integral con respecto de x (entre paréntesis rectos), introducimos el cambio de variable $t = (1 - \rho^2)^{-1/2}((x - a_1)/\sigma_1 - \rho(y - a_2)/\sigma_2)$. Calculando t^2 , y sustituyendo en la mencionada integral, obtenemos

$$\int_{-\infty}^{\infty} (x - a_1)p(x, y) dx = \frac{\rho\sigma_1}{\sigma_2^2\sqrt{2\pi}}(y - a_2)e^{-(y-a_2)^2/(2\sigma_2^2)}.$$

Entonces, sustituyendo el resultado, obtenemos

$$\mathbf{cov}(X, Y) = \frac{\rho\sigma_1}{\sigma_2^2\sqrt{2\pi}} \int_{-\infty}^{\infty} (y - a_2)^2 e^{-(y-a_2)^2/(2\sigma_2^2)} = \rho\sigma_1\sigma_2,$$

y en conclusión, tenemos

$$\rho(X, Y) = \frac{\mathbf{cov}(X, Y)}{\sigma_X\sigma_Y} = \rho.$$

Queda claro entonces el sentido probabilístico del parámetro ρ : es igual al coeficiente de correlación entre las variables aleatorias X e Y .

Como vimos en el ejemplo 3.16, si en la fórmula de la densidad $p(x, y)$ de una variable aleatoria normal bidimensional tenemos $\rho = 0$, entonces las variables X e Y resultan independientes. De esta forma, si bien para dos variables aleatorias arbitrarias su no correlación no implica su independencia, si el vector aleatorio (X, Y) tiene distribución normal, la no correlación entre las variables X e Y es equivalente a su independencia.

Consideremos ahora un vector aleatorio $X = (X_1, \dots, X_n)$. Definimos su esperanza matemática como $\mathbf{E} X = (\mathbf{E} X_1, \dots, \mathbf{E} X_n)$, cuando todas las esperanzas existen. Consideremos la matriz

$$B = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

donde $b_{ij} = \mathbf{cov}(X_i, X_j)$ ($i, j = 1, \dots, n$), y suponemos que todos los momentos existen.

La matriz B se denomina *matriz de covarianza*, o también, *matriz de segundos momentos*. Observemos que la matriz B es simétrica, es decir $b_{ij} = b_{ji}$ para todo par i, j , y que los elementos en su diagonal son las varianzas de las variables aleatorias X_1, \dots, X_n .

Si estas varianzas son todas positivas (es decir, si $b_{ii} > 0$ para cada $i = 1, \dots, n$), podemos definir los coeficientes de correlación $\rho_{ij} = \rho(X_i, X_j) = b_{ij}/(\sqrt{b_{ii}}\sqrt{b_{jj}})$. La matriz

$$\begin{bmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & & \vdots \\ \rho_{n1} & \cdots & \rho_{nn} \end{bmatrix}$$

se denomina *matriz de correlación*. Esta matriz también es simétrica, y sus elementos en la diagonal valen 1.

Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con esperanza $a = \mathbf{E} X = (a_1, \dots, a_n)$, donde $a_k = \mathbf{E} X_k$ ($k = 1, \dots, n$) y matriz de covarianza B no singular, de forma que $\det(B) \neq 0$. Demostremos que $\det(B) > 0$.

En efecto, consideremos la forma cuadrática $\mathbf{E} \left(\sum_{k=1}^n t_k (X_k - a_k) \right)^2 = \sum_{i,j=1}^n t_i t_j b_{ij}$. Para todos los valores de t_1, \dots, t_n esta forma cuadrática es no negativa, basándonos en un teorema de álgebra lineal obtenemos que $\det(B) \geq 0$. Como supusimos que $\det(B) \neq 0$, resulta $\det(B) > 0$.

Ejemplo 4.13. Consideremos, como en el ejemplo 3.15, un vector aleatorio $X = (X_1, \dots, X_n)$, con distribución normal n -dimensional, y densidad dada por

$$p(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} e^{-\frac{1}{2}(x-a)B^{-1}(x-a)'}$$

donde $x = (x_1, \dots, x_n)$ y $a = (a_1, \dots, a_n)$ son vectores fila de números reales; B es una matriz de dimensión $n \times n$, definida positiva no singular y simétrica, B^{-1} es la matriz inversa de la matriz B , y x' denota el vector traspuesto de x .

Es posible demostrar, como se hizo para el caso $n = 2$, que para cada $k = 1, \dots, n$ las variables aleatorias X_k tienen distribución normal con parámetros (a_k, σ_k) , y que covarianza entre las variables aleatorias X_i y X_j es b_{ij} ($i, j = 1, \dots, n$). (Para ésto se utiliza una fórmula análoga a (4.10) para vectores con n coordenadas y se procede como en el ejemplo (4.12).) Esto da sentido probabilístico al vector a y a la matriz B que aparecen en la fórmula de la densidad de X : el vector a es la esperanza del vector aleatorio X ; la matriz B , su matriz de covarianzas.

4.6. Ejercicios

1. Sea X una variable aleatoria con distribución uniforme en el intervalo (a, b) . Hallar $\mathbf{E} X$, $\mathbf{var} X$ y $\mathbf{E}(X^3)$.
2. Sea X una variable aleatoria con distribución discreta, que toma, con probabilidad $1/n$, cada uno de los valores $1, 2, \dots, n$. Calcular $\mathbf{var} X$.
3. Sea X una variable aleatoria con distribución normal con parámetros $(0, 1)$. Calcular $\mathbf{E}(X^k)$ para $k \geq 3$ y la función generatriz de momentos.

4. Consideremos una variable aleatoria X con distribución exponencial, con parámetro $\alpha > 0$. Calcular $\mathbf{E} X$, $\mathbf{var} X$ y $\mathbf{E}(X^3)$.

5. Sea X una variable aleatoria que toma únicamente valores enteros no negativos. Demostrar que $\mathbf{E} X = \sum_{n=1}^{\infty} \mathbf{P}(X \geq n)$.

6. Sea X una variable aleatoria arbitraria. Demostrar que

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

7. Sea X una variable aleatoria para la cual existe esperanza matemática $\mathbf{E} X$. Demostrar que la función de distribución $F(x)$ de esta variable aleatoria cumple que $\lim_{x \rightarrow \infty} x(1 - F(x)) = 0$.

8. Sea X una variable aleatoria no negativa para la cual existe esperanza matemática $\mathbf{E} X$. Demostrar que $\mathbf{E} X = \int_0^{\infty} (1 - F(x)) dx$, donde $F(x)$ es la función de distribución de la variable aleatoria X .

9. Sea X una variable aleatoria para la cual existe esperanza matemática $\mathbf{E} X$. Demostrar que

$$\mathbf{E} X = - \int_{-\infty}^0 F(x) dx + \int_0^{\infty} (1 - F(x)) dx,$$

donde $F(x)$ es la función de distribución de la variable aleatoria X .

10. Sea X una variable aleatoria. Demostrar que si $\mathbf{E}|X|^r < \infty$ para algún $r > 0$, entonces

$$\mathbf{E}|X|^r = r \int_0^{\infty} \mathbf{P}(|X| \geq x) x^{r-1} dx.$$

11. Sea X e Y dos variables aleatorias con funciones de distribución $F(x)$ y $G(x)$ respectivamente, para las cuales existe esperanza matemática. Demostrar que

$$\mathbf{E}(X - Y) = \int_{-\infty}^{\infty} (G(x) - F(x)) dx.$$

12. La variable aleatoria X tiene *distribución de Laplace*, si su densidad es

$$p(x) = \frac{1}{2b} e^{-|x-a|/b},$$

donde $b > 0$ y a son constantes. Hallar $\mathbf{E} X$ y $\mathbf{var} X$.

13. La densidad de la magnitud de la velocidad absoluta de una molécula tiene la forma

$$p(x) = \frac{4x^2}{\alpha^3 \sqrt{\pi}} e^{-x^2/\alpha^2}, \quad \text{si } x > 0,$$

y $p(x) = 0$ si $x \leq 0$ (*distribución de Maxwell*). Hallar la velocidad media de una molécula y su varianza.

14. Consideremos una variable aleatoria X con distribución normal con parámetros (a, σ) . Demostrar que la variable aleatoria $Y = e^X$ tiene densidad dada por

$$p(y) = \frac{1}{\sigma y \sqrt{2\pi}} e^{-(\ln y - a)^2 / (2\sigma^2)}, \quad \text{si } x > 0,$$

y $p(x) = 0$ si $x \leq 0$ (la distribución con esta densidad se denomina *log-normal*). Hallar $\mathbf{E} Y$ y $\mathbf{var} Y$ en el caso $a = 0, \sigma = 1$.

15. Una persona quiere abrir una puerta, y tiene n llaves de las cuales solo una corresponde a la cerradura. La persona va eligiendo las llaves al azar y probando abrir la puerta. Calcular la esperanza matemática y la varianza del número de intentos en cada uno de los dos siguientes casos: (a) la persona elige una llave nueva cada vez, (b) la persona elige cada vez entre las n llaves.

16. Sean X, Y dos variables aleatorias para las cuales existe la varianza. Demostrar que $\mathbf{var}(X + Y) \leq 2(\mathbf{var} X + \mathbf{var} Y)$.

17. Sean X, Y dos variables aleatorias para las cuales existen las varianzas $\mathbf{var} X = \sigma_X^2, \mathbf{var} Y = \sigma_Y^2$. Notemos $\mathbf{var}(X + Y) = \sigma_{X+Y}^2$. Demostrar que $\sigma_{X+Y} \leq \sigma_X + \sigma_Y$.

18. Sea $f(x)$ una función definida en la recta real, no negativa y no decreciente, y sea X una variable aleatoria tal que existe $\mathbf{E} f(X)$. Demostrar la desigualdad $\mathbf{P}(X \geq x) \leq \frac{1}{f(x)} \mathbf{E} f(X)$ para todo x real.

19. Construir la densidad de una distribución de una variable aleatoria X tal que $\mathbf{E}(X^2) < \infty$, y $\mathbf{E}|X|^3 = \infty$.

20. Construir la densidad de una distribución de una variable aleatoria X tal que:

(a) $\mathbf{E}|X|^3 = \infty$, y $\mathbf{E}|X|^{2+\delta} < \infty$ para cualquier δ positivo, $\delta < 1$;

(b) $\mathbf{E}|X|^{2+\delta} = \infty$, y $\mathbf{E}(X^2) < \infty$ para cualquier $\delta > 0$.

21. Calcular la función generatriz de momentos de una variable aleatoria X con distribución exponencial de parámetro $\alpha = 1$. Utilizando este resultado, calcular los momentos $E(X^k)$ para k natural, $k \geq 1$.

22. Calcular la función generatriz de momentos de una variable aleatoria X con distribución de Poisson de parámetro λ .

23. Calcular la función generatriz de momentos de una variable aleatoria X con distribución binomial de parámetros (n, p) .

24. Calcular la mediana de una variable aleatoria con distribución exponencial de parámetro $\alpha = 1$.

25. Consideremos una variable aleatoria X con función de distribución X dada por

$$F(x) = 1 - e^{-x^b/c}, \quad \text{si } x > 0,$$

y $p(x) = 0$ si $x \leq 0$, donde b y c son constantes positivas (*distribución de Weibull*). Hallar la mediana y los cuantiles de orden q de la variable aleatoria X .

26. El vector aleatorio (X, Y) tiene densidad dada por

$$p(x, y) = \begin{cases} \frac{1}{3}(x + y), & \text{si } 0 \leq x \leq 2, 0 \leq y \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

Calcular $\mathbf{E}X$ y $\mathbf{E}Y$.

27. En el problema 26 calcular $\mathbf{var}(3X - 2Y + 5)$.

28. Sean X, Y variables aleatorias, tales que $\mathbf{E}X = 1$, $\mathbf{E}Y = 2$, $\mathbf{var} X = 1$, $\mathbf{var} Y = 4$. Sea $Z = 3X - Y + 9$. Hallar $\mathbf{E}Z$ y $\mathbf{var} Z$ en cada uno de los siguientes casos: (a) X e Y son independientes. (b) X e Y son no correlacionadas. (c) El coeficiente de correlación $\rho(X, Y) = 0,6$.

29. Sean X, Y variables aleatorias, tales que $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathbf{var} X = \mathbf{var} Y = 1$, $\rho(X, Y) = r$. Calcular: (a) $\mathbf{var}(X - rY)$; (b) el coeficiente de correlación entre $X - rY$ e Y .

Capítulo 5

Distintos tipos de convergencia en teoría de la probabilidad. Ley de los grandes números

5.1. Distintos tipos de convergencia en teoría de la probabilidad.

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$ y variables aleatorias X, X_1, X_2, \dots definidas en este espacio.

Definición 5.1. *Decimos que la sucesión de variables aleatorias $\{X_n\}$ converge en probabilidad a la variable aleatoria X , y escribimos $X_n \xrightarrow{\mathbf{P}} X$, si para todo $\varepsilon > 0$ se verifica*

$$\mathbf{P}(|X_n - X| \geq \varepsilon) \rightarrow 0,$$

cuando $n \rightarrow \infty$.

Es claro que esta definición equivale a la siguiente: Para todo $\varepsilon > 0$, se verifica

$$\mathbf{P}(|X_n - X| < \varepsilon) \rightarrow 1,$$

cuando $n \rightarrow \infty$.

Definición 5.2. Decimos que la sucesión de variables aleatorias $\{X_n\}$ converge casi seguramente a la variable aleatoria X , y escribimos $X_n \rightarrow X$ c.s., si $X_n(\omega) \rightarrow X(\omega)$ para todos los puntos $\omega \in \Omega$, con excepción de un conjunto de probabilidad nula. En otras palabras, $X_n \rightarrow X$ c.s. si se verifica $\mathbf{P}(\{\omega: X_n(\omega) \rightarrow X(\omega)\}) = 1$.

En análisis real, la convergencia en probabilidad corresponde a la *convergencia en medida*; la convergencia casi segura, a la *convergencia en casi todo punto*. De forma que, como la convergencia en casi todo punto implica la convergencia en medida, la convergencia casi segura implica la convergencia en probabilidad. Veamos una demostración directa de este resultado, sin apelar a nociones del análisis real.

Supongamos entonces que $X_n \rightarrow X$ c.s. Esto significa que dado $\varepsilon > 0$, para cada $\omega \in \Omega$, con excepción de un suceso de probabilidad nula, existe $n = n(\varepsilon, \omega)$ tal que para todo $k \geq n(\varepsilon, \omega)$, tenemos $|X_k(\omega) - X(\omega)| < \varepsilon$. Es decir, dado $\varepsilon > 0$, tenemos

$$\mathbf{P} \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{|X_k(\omega) - X(\omega)| < \varepsilon\} \right) = 1.$$

Tomando complemento se obtiene que, dado $\varepsilon > 0$, tenemos

$$\mathbf{P} \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|X_k(\omega) - X(\omega)| \geq \varepsilon\} \right) = 0. \quad (5.1)$$

La sucesión de conjuntos $\mathbf{E}_n = \bigcup_{k=n}^{\infty} \{|X_k(\omega) - X(\omega)| \geq \varepsilon\}$ cumple $\mathbf{E}_1 \supset \mathbf{E}_2 \supset \dots$. Dado que se verifica (5.1), por la propiedad 8 (sección 1.3) obtenemos $\lim_{n \rightarrow \infty} \mathbf{P}(\mathbf{E}_n) = \mathbf{P}(\bigcap_{n=1}^{\infty} \mathbf{E}_n) = 0$. De aquí se obtiene, que

$$\mathbf{P}(|X_n - X| \geq \varepsilon) \leq \mathbf{P} \left(\bigcup_{k=n}^{\infty} |X_k - X| \geq \varepsilon \right) = \mathbf{P}(\mathbf{E}_n) \rightarrow 0,$$

si $n \rightarrow \infty$, para $\varepsilon > 0$ arbitrario. En consecuencia $X_n \xrightarrow{\mathbf{P}} X$, concluyendo la demostración.

Definición 5.3. Supongamos que las variables aleatorias X, X_1, X_2, \dots tienen momento finito de orden $r \geq 1$. Decimos que la sucesión de variables aleatorias $\{X_n\}$ converge en media de orden r , o más brevemente, converge en r -media a la variable aleatoria X , si $\mathbf{E}|X_n - X|^r \rightarrow 0$ cuando $n \rightarrow \infty$.

No existe una notación estándar para este tipo de convergencia, y aquí utilizaremos la notación $X_n \rightarrow X$ (r -media). Si $r = 2$ se dice que la sucesión converge *en media cuadrática*, si $r = 1$ que *converge en media*.

Demostremos que si $X_n \rightarrow X$ (r -media) para $r \geq 1$, entonces $X_n \xrightarrow{\mathbf{P}} X$. Aplicando la desigualdad de Chebishev, tenemos

$$\mathbf{P}(|X_n - X| \geq \varepsilon) = \mathbf{P}(|X_n - X|^r \geq \varepsilon^r) \leq \frac{1}{\varepsilon^r} \mathbf{E}|X_n - X|^r,$$

para cualquier $\varepsilon > 0$. Si $\mathbf{E}|X_n - X|^r \rightarrow 0$, el lado derecho en la desigualdad anterior tiende a cero, y por lo tanto $\mathbf{P}(|X_n - X| \geq \varepsilon) \rightarrow 0$ para todo $\varepsilon > 0$, es decir, $X_n \xrightarrow{\mathbf{P}} X$.

Definición 5.4. Consideremos las variables aleatorias X, X_1, X_2, \dots , y sus funciones de distribución $F(x), F_1(x), F_2(x) \dots$. Decimos que la sucesión de variables aleatorias $\{X_n\}$ converge en distribución a la variable aleatoria X , si $F_n(x) \rightarrow F(x)$ en cada punto x de continuidad de la función $F(x)$.

Además de decirse *convergencia en distribución*, se utiliza el término *convergencia débil de distribuciones*. Tampoco existe una notación establecida para la convergencia en distribución, siendo las más comunes

$$X_n \xrightarrow{d} X, \quad \text{y} \quad X_n \xrightarrow{\mathcal{L}} X.$$

Proposición 5.1. Consideremos las variables aleatorias $X, X_1, X_2 \dots$, con funciones de distribución $F(x), F_1(x), F_2(x) \dots$. Si $X_n \xrightarrow{\mathbf{P}} X$, entonces $X_n \xrightarrow{d} X$.

La demostración se basa en el siguiente resultado.

Lema 5.1. Supongamos que $X_n \xrightarrow{\mathbf{P}} X$, y sean $x' < x < x''$ números reales arbitrarios. Entonces

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x'').$$

Demostración. Tenemos

$$\begin{aligned} F(x') &= \mathbf{P}(X \leq x') \\ &= \mathbf{P}(X \leq x', X_n - X \geq x - x') + \mathbf{P}(X \leq x', X_n - X < x - x'), \end{aligned}$$

dado que $\mathbf{P}(\mathbf{A}) = \mathbf{P}(\mathbf{A}\mathbf{B}) + \mathbf{P}(\mathbf{A}\bar{\mathbf{B}})$ para sucesos \mathbf{A} y \mathbf{B} arbitrarios. Además, $\mathbf{P}(\mathbf{A}\mathbf{B}) \leq \mathbf{P}(\mathbf{B})$, y por ésto

$$F(x') \leq \mathbf{P}(X_n - X \geq x - x') + \mathbf{P}(X_n \leq x), \quad (5.2)$$

dado que, el producto de los sucesos $X \leq x'$ y $X_n - X < x - x'$ implica el suceso $X_n \leq x$. Como $x' < x < x''$, tenemos $x - x' > 0$, y el primer sumando en (5.2) tiene límite nulo cuando $n \rightarrow \infty$, porque $X_n \xrightarrow{\mathbf{P}} X$. Tomando límite inferior a ambos lados en (5.2), obtenemos

$$F(x') \leq \liminf_{n \rightarrow \infty} F_n(x).$$

Un razonamiento análogo conduce a la desigualdad

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x'').$$

(Recomendamos al lector llevar a cabo este razonamiento.) Esto concluye la demostración. \square

Demostración de la proposición 5.1. Consideremos un punto de continuidad x de la función $F(x)$. Para cada $k = 1, 2, \dots$, introducimos $x'_k = x - 1/k$, $x''_k = x + 1/k$. Es claro, que se verifica $x'_k < x < x''_k$. Por ésto, aplicando el lema 5.1 (con k fijo), tenemos

$$F(x'_k) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x''_k).$$

Es claro además, que $\lim_k x'_k = \lim_k x''_k = x$ ($k \rightarrow \infty$). Como $F(x)$ es continua en el punto x , al tomar límite en la fórmula anterior, si $k \rightarrow \infty$, obtenemos

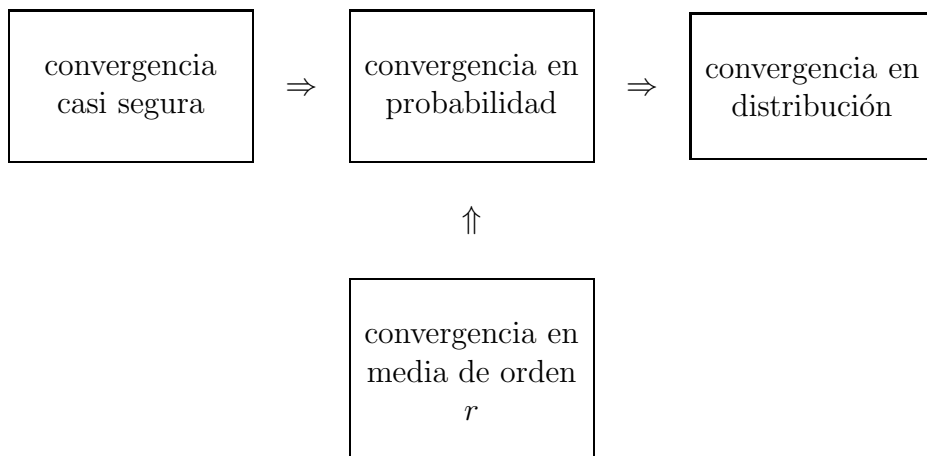
$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x),$$

lo que implica, que

$$\liminf_{n \rightarrow \infty} F_n(x) = \limsup_{n \rightarrow \infty} F_n(x) = F(x).$$

Existe entonces el $\lim_n F_n(x) = F(x)$ para todos los puntos de continuidad, concluyendo la demostración. \square

Como conclusión de las proposiciones demostradas, hemos obtenido que son válidas las siguientes implicaciones entre los distintos tipos de convergencia introducidos:



Observemos que esta tabla incluye todas las relaciones posibles entre los tipos de convergencia considerados. Sin supuestos adicionales, no es posible obtener más relaciones.

5.2. Ley de los grandes números

Se denomina *ley de los grandes números* a cualquier proposición que establece, bajo determinadas condiciones, la convergencia en probabilidad o casi segura de los promedios aritméticos de una cantidad creciente a infinito de variables aleatorias. Si tenemos convergencia en probabilidad, decimos *ley débil de los grandes números*, si tenemos convergencia casi segura, *ley fuerte de los grandes números*. En esta sección, estudiaremos diversas formas de la ley débil de los grandes números¹.

Teorema 5.1 (Chebishev). *Consideremos una sucesión de variables aleatorias X_1, X_2, \dots , independientes dos a dos, y con esperanzas matemáticas a_1, a_2, \dots . Supongamos que se cumple la condición*

$$\text{var } X_n \leq C \text{ para cada } n = 1, 2, \dots,$$

¹En la sección 9.5 estudiamos una ley fuerte de los grandes números.

donde C es una constante arbitraria. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \xrightarrow{\mathbf{P}} 0. \quad (5.3)$$

Es importante el caso particular en el que las esperanzas de las variables aleatorias son todas iguales, digamos $\mathbf{E} X_n = a$ para todo $n = 1, 2, \dots$. En esta situación tenemos $\sum_{i=1}^n a_i/n = a$, y la convergencia en (5.3) se transforma en

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

Demostración. Consideremos para cada $n = 1, 2, \dots$ la variable aleatoria $Z_n = \sum_{i=1}^n X_i/n$. Es claro que $\mathbf{E} Z_n = \sum_{i=1}^n a_i/n$. Como las variables aleatorias son independientes dos a dos, tenemos $\mathbf{var} \sum_{i=1}^n X_i = \sum_{i=1}^n \mathbf{var} X_i \leq nC$, en vista de la propiedad 5. Aplicamos ahora la desigualdad de Chebishev (4.21), para obtener

$$\begin{aligned} \mathbf{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \right| \geq \varepsilon \right) &= \mathbf{P} (|Z_n - \mathbf{E} Z_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbf{var} Z_n \\ &= \frac{1}{n^2 \varepsilon^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0, \end{aligned} \quad (5.4)$$

si $n \rightarrow \infty$, dado que $\sum_{i=1}^n \mathbf{var} X_i \leq nC$. Esto concluye la demostración. \square

El teorema de Chebishev da fundamento a la *regla de la media aritmética*, utilizada en el procesamiento de resultados de mediciones: para la estimación del valor de una constante física a , desconocida, mediante los resultados de n mediciones de su magnitud, se recomienda tomar el promedio aritmético de estas mediciones. Veamos la fundamentación. Supongamos que X_1, \dots, X_n son los resultados de las n mediciones de esta constante a . Por cuanto las mediciones habitualmente se acompañan de errores, y no podemos predecir el resultado de la medición siguiente, consideramos que el resultado de la i -ésima ($i = 1, \dots, n$) medición de la constante a es una variable aleatoria $X_i = a + \Delta_i$, donde Δ_i es el error que se comete en la i -ésima medición. Suponemos que

$$\mathbf{E} \Delta_i = 0, \quad \mathbf{var} \Delta_i = \sigma^2 \quad (i = 1, \dots, n), \quad (5.5)$$

donde el valor de σ^2 puede ser desconocido. Las condiciones en (5.5) son equivalentes a la condiciones

$$\mathbf{E} X_i = a, \quad \mathbf{var} X_i = \sigma^2 \quad (i = 1, \dots, n). \quad (5.6)$$

La primer condición en (5.6) se interpreta como la ausencia de error sistemático en las mediciones, la segunda significa que todas las mediciones se llevan a cabo con la misma precisión. Si se supone además que las variables aleatorias X_1, \dots, X_n son independientes dos a dos, es aplicable el teorema de Chebishev con $a_1 = \dots = a_n = a$ y la constante $C = \sigma^2$, de acuerdo al cual obtenemos

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

En consecuencia, el promedio aritmético de los resultados de las mediciones de una constante física a converge en probabilidad al valor de esta constante; en la terminología de la estadística matemática se dice que el promedio aritmético de las mediciones es un *estimador consistente* de la constante desconocida a .

Veamos otro corolario del teorema de Chebishev, demostrando, que de este teorema se deduce el teorema de Bernoulli de la sección 2.4.

Consideremos entonces una serie de n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso), y probabilidad de éxito igual a p en cada experimento ($0 < p < 1$). Sea μ la cantidad de éxitos en n experimentos. Veamos como se formula el teorema de Bernoulli por medio de variables aleatorias. Para ésto, sea X_1, X_2, \dots una sucesión de variables aleatorias, cada una de las cuales toma el valor 1 con probabilidad p (si ocurre un éxito) y el valor 0 con probabilidad $q = 1 - p$ (si ocurre un fracaso). Tenemos $\mathbf{E} X_i = p$, $\mathbf{var} X_i = pq \leq 1$ para cada $i = 1, 2, \dots$. Además $\mu = \sum_{i=1}^n X_i$, porque la suma contiene tantos sumandos iguales a uno como éxitos ocurren en los primeros n experimentos, siendo nulos los sumandos restantes. Las variables aleatorias X_1, X_2, \dots son independientes dos a dos y cumplen la condición $\mathbf{var} X_i \leq 1$ ($i = 1, 2, \dots$), por lo que es aplicable el teorema de Chebishev, y de (5.3) se obtiene $\mu/n \xrightarrow{\mathbf{P}} p$. La última afirmación significa que la frecuencia de éxitos en n experimentos, converge en probabilidad a p , la probabilidad de éxito en un experimento. Éste es el contenido del teorema de Bernoulli del capítulo 2 (página 50).

Teorema 5.2 (Markov). *Consideremos una sucesión de variables aleatorias X_1, X_2, \dots con esperanzas matemáticas a_1, a_2, \dots . Supongamos que se cumple la condición*

$$\frac{1}{n^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0, \quad (5.7)$$

si $n \rightarrow \infty$. Entonces tiene lugar la convergencia en (5.3), es decir

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n a_i \xrightarrow{\mathbf{P}} 0.$$

Demostración. Al igual que en la demostración del teorema de Chebishev, aplicamos la desigualdad de Chebishev con la misma elección de la variable aleatoria Z_n . Tenemos, como antes, $\mathbf{E} Z_n = \sum_{i=1}^n a_i/n$. Por la condición (5.7), tenemos

$$\mathbf{var} Z_n = \frac{1}{n^2} \mathbf{var} \sum_{i=1}^n X_i \rightarrow 0 \text{ si } n \rightarrow \infty.$$

Esto muestra que la acotación y el límite en la fórmula (5.4) tienen lugar, concluyendo la demostración. \square

El teorema de Markov es una generalización del teorema de Chebishev. La tesis en ambos teoremas es la misma, pero la hipótesis en el teorema de Markov es más general que en teorema de Chebishev. Si se cumplen las hipótesis del teorema de Chebishev tenemos $\mathbf{var} \sum_{i=1}^n X_i = \sum_{i=1}^n \mathbf{var} X_i \leq nC$, dada la independencia dos a dos de las variables aleatorias X_1, X_2, \dots y la condición (5.1), por lo que la condición (5.7) se cumple de manera evidente.

Es importante observar que en el teorema de Markov no figura supuesto alguno sobre la *independencia* de las variables aleatorias consideradas. Ésto permite su aplicación para la demostración de la ley de los grandes números (es decir, la demostración de la afirmación (5.3)) para sucesiones de variables aleatorias dependientes.

Veamos como obtener, con la ayuda del teorema de Markov, una ley de los grandes números para una sucesión estacionaria.

Decimos que una sucesión de variables aleatorias X_1, X_2, \dots es una *sucesión estacionaria* (más precisamente, una *sucesión débilmente estacionaria*), si la esperanza de las variables aleatorias $\mathbf{E} X_n$ es constante para

todo n , existen los momentos de segundo orden para todo n , y además, la esperanza $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$.

Ejemplo 5.1. Consideremos una sucesión de variables aleatorias $\{X_n\}$ no correlacionadas dos a dos, y tal que $\mathbf{E} X_n = 0$, $\mathbf{var} X_n = 1$ para todo $n = 1, 2, \dots$. Veamos que es estacionaria.

Sea $\rho(X_n, X_m)$ el coeficiente de correlación entre X_n y X_m , es decir

$$\rho(X_n, X_m) = \frac{\mathbf{E}(X_n X_m) - \mathbf{E} X_n \mathbf{E} X_m}{\sqrt{\mathbf{var} X_n} \sqrt{\mathbf{var} X_m}}.$$

En vista de nuestros supuestos $\rho(X_n, X_m) = \mathbf{E}(X_n X_m) = 0$ si $n \neq m$, y en el caso $m = n$ obtenemos $\rho(X_n, X_n) = \mathbf{E} X_n^2 = 1$. En consecuencia

$$\mathbf{E}(X_n X_m) = \begin{cases} 0, & \text{si } n - m \neq 0, \\ 1, & \text{si } n - m = 0, \end{cases}$$

es decir, la esperanza $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$. De aquí concluimos que la sucesión de variables aleatorias considerada es estacionaria.

Ejemplo 5.2. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas. Supongamos que existen $\mathbf{E} X_n = a$ y $\mathbf{var} X_n = \sigma^2$. Esta sucesión de variables aleatorias es estacionaria.

En efecto, si $m = n$, tenemos $\mathbf{E} X_n^2 = \mathbf{var} X_n + (\mathbf{E} X_n)^2 = \sigma^2 + a^2$. Si $m \neq n$, como la correlación es nula, es nula la covarianza, y tenemos $\mathbf{E} X_n X_m = \mathbf{E} X_n \mathbf{E} X_m = a^2$. En conclusión

$$\mathbf{E}(X_n X_m) = \begin{cases} a^2, & \text{si } n - m \neq 0, \\ \sigma^2 + a^2, & \text{si } n - m = 0, \end{cases}$$

y $\mathbf{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$.

Teorema 5.3. *Consideremos una sucesión estacionaria de variables aleatorias X_1, X_2, \dots , con $\mathbf{E} X_n = a$. Supongamos que $\rho(X_n, X_m) \rightarrow 0$ cuando $|n - m| \rightarrow \infty$, donde $\rho(X_n, X_m)$ es el coeficiente de correlación entre las variables aleatorias X_n y X_m . Entonces*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} a.$$

Demostración. Es suficiente demostrar que nuestras hipótesis implican la condición (5.7) del teorema de Markov.

Como la sucesión es estacionaria $\mathbf{E} X_n^2$ no depende de n , por lo que $\mathbf{var} X_n = \mathbf{E} X_n^2 - (\mathbf{E} X_n)^2 = \mathbf{E} X_n^2 - a^2$ tampoco depende de n . Pongamos entonces $\sigma^2 = \mathbf{var} X_n$. Por la proposición 4.2, tenemos

$$\begin{aligned} \mathbf{var} \sum_{i=1}^n X_i &= \sum_{i=1}^n \mathbf{var} X_i + 2 \sum_{1 \leq i < k \leq n} \mathbf{cov}(X_i, X_k) \\ &= n\sigma^2 + 2\sigma^2 \sum_{1 \leq i < k \leq n} \rho(X_i, X_k) = n\sigma^2 + 2\sigma^2(T_1 + T_2) \end{aligned}$$

donde

$$T_1 = \sum_{\substack{1 \leq i < k \leq n \\ |i-k| < M}} \rho(X_i, X_k), \quad T_2 = \sum_{\substack{1 \leq i < k \leq n \\ |i-k| \geq M}} \rho(X_i, X_k),$$

y M es una constante positiva.

Sea $\varepsilon > 0$ arbitrario. Existe M tal que se verifica la desigualdad $|\rho(X_i, X_k)| < \varepsilon$ si $|i - k| \geq M$, dado que $|\rho(X_i, X_k)| \rightarrow 0$ si $|i - k| \rightarrow \infty$. Entonces $|T_2| \leq \varepsilon n^2$. Para acotar la suma T_1 aplicamos la desigualdad $|\rho(X_n, X_m)| \leq 1$ (proposición 4.3), y obtenemos $|T_1| \leq (2M + 1)n$. En conclusión,

$$\mathbf{var} \sum_{i=1}^n X_i \leq n\sigma^2 + 2\sigma^2((2M + 1)n + \varepsilon n^2)$$

de donde, dividiendo por n^2 , obtenemos

$$\frac{1}{n^2} \mathbf{var} \sum_{i=1}^n X_i \leq \frac{\sigma^2}{n} + 2\sigma^2 \left(\frac{2M + 1}{n} + \varepsilon \right).$$

Como ε es arbitrario, esta última acotación conduce a la condición (5.7) en el teorema de Markov, lo que concluye la demostración. \square

Aplicando la desigualdad de Chebishev es posible obtener una demostración del *teorema de Weierstrass* del análisis real, de acuerdo al cual, para cualquier función continua $f(x)$ definida en el intervalo $[0, 1]$, existe una sucesión de polinomios $\{P_n(x)\}$ tal que $P_n(x) \rightarrow f(x)$ uniformemente en el intervalo $[0, 1]$, cuando $n \rightarrow \infty$.

Teorema 5.4 (Bernstein). *Consideremos una función continua $f(x)$ definida en el intervalo $[0, 1]$. Sea*

$$B_n(x; f) = \sum_{k=0}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k}.$$

Entonces $B_n(x; f) \rightarrow f(x)$ uniformemente, en el intervalo $[0, 1]$, cuando $n \rightarrow \infty$. Las funciones $B_n(x; f)$ son polinomios de grado n , que se denominan polinomios de Bernstein.

Demostración. Consideremos la diferencia

$$f(x) - B_n(x; f) = \sum_{k=0}^n [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k} = T_1 + T_2,$$

donde

$$T_1 = \sum_{k: |\frac{k}{n} - x| < \delta} [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k},$$

$$T_2 = \sum_{k: |\frac{k}{n} - x| \geq \delta} [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k}.$$

Primero obtenemos una acotación para el término T_1 . Como $f(x)$ es una función continua en el intervalo $[0, 1]$, dado $\varepsilon > 0$ existe $\delta > 0$ tal que $|x - y| < \delta$ implica $|f(x) - f(y)| < \varepsilon$. Entonces

$$\begin{aligned} |T_1| &\leq \sum_{k: |\frac{k}{n} - x| < \delta} |f(x) - f(k/n)| \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \varepsilon \sum_{k: |\frac{k}{n} - x| < \delta} \binom{n}{k} x^k (1-x)^{n-k} \leq \varepsilon (x + (1-x))^n = \varepsilon. \end{aligned} \quad (5.8)$$

La acotación del término T_2 , se obtiene aplicando la desigualdad de Chebishev en un esquema de Bernoulli. Consideremos entonces una serie de n experimentos independientes con dos resultados posibles cada uno: éxito con probabilidad x , y fracaso con probabilidad $1 - x$. Si μ designa la cantidad de éxitos en n experimentos, dado $\delta > 0$, tenemos

$$\mathbf{P}\left(\left|\frac{\mu}{n} - x\right| \geq \delta\right) = \sum_{k: |\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1-x)^{n-k}.$$

Como $\mathbf{E}(\mu/n) = x$, $\mathbf{var}(\mu/n) = x(1-x)/n$, y $|f(x)| \leq C$ donde C es una constante, aplicando la desigualdad de Chebishev, tenemos

$$\begin{aligned} |T_2| &\leq 2C \sum_{k: |\frac{k}{n} - x| \geq \delta} \binom{n}{k} x^k (1-x)^{n-k} \\ &= 2C \mathbf{P} \left(\left| \frac{\mu}{n} - x \right| \geq \delta \right) \leq 2C \frac{x(1-x)}{n\delta^2} \leq \frac{C}{2n\delta^2} < \varepsilon \end{aligned} \quad (5.9)$$

(donde usamos que $x(1-x) \leq 1/4$), si n es suficientemente grande. En conclusión, como las acotaciones en (5.8) y (5.9) son independientes de x tenemos que, para n suficientemente grande, vale

$$\sup_{0 \leq x \leq 1} |f(x) - B_n(x; f)| \leq |T_1| + |T_2| < 2\varepsilon,$$

concluyendo la demostración. \square

Como conclusión de esta sección presentamos diferentes condiciones para la validez de la ley débil de los grandes números. Estamos particularmente interesados en debilitar las condiciones en los momentos de las variables aleatorias (como por ejemplo la condición (5.1) en el teorema de Chebishev, o la condición (5.7) en el teorema de Markov), mas aún en obtener resultados sin condiciones de momentos. Con este fin asumiremos, por una parte, que las variables aleatorias X_1, X_2, \dots están *idénticamente distribuidas*, y por otra que son independientes dos a dos.

Teorema 5.5. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas, con función de distribución $V(x)$. Supongamos que se verifican las dos siguientes condiciones:*

$$n \mathbf{P}(|X_1| \geq n) \rightarrow 0, \quad (5.10)$$

$$\int_{\{|x| < n\}} x dV(x) \rightarrow 0, \quad (5.11)$$

cuando $n \rightarrow \infty$. Entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} 0. \quad (5.12)$$

Antes de comenzar la demostración observemos que, si la condición de independencia dos a dos se sustituye por la condición de independencia

mutua de las variables X_1, \dots, X_n , para cada $n = 1, 2, \dots$, entonces las condiciones (5.10) y (5.11) no solo son suficientes sino también necesarias. Este último resultado es consecuencia de un teorema más general, obtenido por Kolmogorov, para sucesiones de variables aleatorias independientes, no necesariamente idénticamente distribuidas, en el que no se asume ninguna condición de existencia de momentos de estas variables aleatorias².

Demostración del teorema 5.5. Consideremos para cada $n = 1, 2, \dots$ las variables aleatorias $X_{ni} = X_i/n$ ($1 \leq i \leq n$). El límite en (5.12), en términos de estas nuevas variables, es

$$\sum_{i=1}^n X_{ni} \xrightarrow{\mathbf{P}} 0. \quad (5.13)$$

Consideremos ahora las variables aleatorias truncadas, definidas mediante

$$\bar{X}_{ni} = \begin{cases} X_{ni}, & \text{si } |X_{ni}| < 1, \\ 0, & \text{si } |X_{ni}| \geq 1. \end{cases}$$

Estas variables aleatorias están acotadas, por lo que existen sus momentos de cualquier orden. Para calcular los dos primeros momentos, observamos que $\bar{X}_{ni} = g(X_i/n)$, donde $g(x) = x\mathbf{1}_{\{|x|<1\}}$. Aplicando entonces la identidad (4.3), tenemos

$$\mathbf{E} \bar{X}_{ni} = \mathbf{E} g(X_i/n) = \int_{-\infty}^{\infty} g(y/n) dV(y) = \frac{1}{n} \int_{\{|y|<n\}} y dV(y), \quad (5.14)$$

$$\mathbf{E} \bar{X}_{ni}^2 = \int_{-\infty}^{\infty} (g(y/n))^2 dV(y) = \frac{1}{n^2} \int_{\{|y|<n\}} y^2 dV(y). \quad (5.15)$$

Demostremos ahora que de la condición (5.10), se obtiene que

$$\frac{1}{n} \int_{\{|x|<n\}} x^2 dV(x) \rightarrow 0 \quad (n \rightarrow \infty). \quad (5.16)$$

En efecto, recordando que $k^2 \leq k(k+1) = 2 \sum_{j=1}^k j$, tenemos

$$\int_{\{|x|<n\}} x^2 dV(x) = \sum_{k=1}^n \int_{\{k-1 \leq |x| < k\}} x^2 dV(x)$$

²Ver por ejemplo §4.4. en V.V. Petrov, Limit Theorems of Probability Theory, (1995)

$$\begin{aligned}
&\leq \sum_{k=1}^n k^2 \mathbf{P}(k-1 \leq |X_1| < k) \\
&\leq \sum_{k=1}^n \left(2 \sum_{j=1}^k j\right) \mathbf{P}(k-1 \leq |X_1| < k) = \\
&= 2 \sum_{j=1}^n j \mathbf{P}(j-1 \leq |X_1| < n) \leq 2 \sum_{j=1}^n j \mathbf{P}(j-1 \leq |X_1|).
\end{aligned}$$

Es sencillo demostrar que si $\{b_j\}$ es una sucesión numérica, tal que $b_j \rightarrow 0$ ($j \rightarrow \infty$), entonces $(\sum_{j=1}^n b_j)/n \rightarrow 0$ ($n \rightarrow \infty$). Si ponemos $b_j = j \mathbf{P}(j-1 \leq |X_1|)$, tenemos

$$b_j = j \mathbf{P}(j-1 \leq |X_1|) = \frac{j}{j-1} (j-1) \mathbf{P}(j-1 \leq |X_1|) \rightarrow 0 \quad (j \rightarrow \infty),$$

en vista de (5.10). Por ésto

$$\frac{1}{n} \int_{\{|x|<n\}} x^2 dV(x) \leq \frac{2}{n} \sum_{j=1}^n b_j \rightarrow 0 \quad (n \rightarrow \infty),$$

y obtenemos (5.16).

Aplicamos ahora la desigualdad de Chebishev. Teniendo en cuenta (5.15) y (5.16), obtenemos

$$\begin{aligned}
\mathbf{P}\left(\left|\sum_{i=1}^n \bar{X}_{ni} - \sum_{i=1}^n \mathbf{E} \bar{X}_{ni}\right| \geq \delta\right) &\leq \frac{1}{\delta^2} \mathbf{var} \sum_{i=1}^n \bar{X}_{ni} = \frac{1}{\delta^2} \sum_{i=1}^n \mathbf{var} \bar{X}_{ni} \\
&\leq \frac{1}{\delta^2} \sum_{i=1}^n \mathbf{E} \bar{X}_{ni}^2 = \frac{1}{\delta^2 n} \int_{\{|x|<n\}} x^2 dV(x) \rightarrow 0,
\end{aligned}$$

para $\delta > 0$ cuando $n \rightarrow \infty$. En los cálculos anteriores utilizamos la propiedad 5 para variables aleatorias independientes dos a dos. De esta forma, obtenemos

$$\sum_{i=1}^n \bar{X}_{ni} - \sum_{i=1}^n \mathbf{E} \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0.$$

En vista de la fórmula (5.14) y la condición (5.11), tenemos

$$\sum_{i=1}^n \mathbf{E} \bar{X}_{ni} = \int_{\{|x|<n\}} x dV(x) \rightarrow 0 \quad (n \rightarrow \infty).$$

Por esto concluimos que $\sum_{i=1}^n \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0$. Resta entonces demostrar, que

$$\sum_{i=1}^n X_{ni} - \sum_{i=1}^n \bar{X}_{ni} \xrightarrow{\mathbf{P}} 0,$$

de donde se obtiene (5.13).

Veamos esta última condición. Para todo $\delta > 0$ tenemos

$$\begin{aligned} \mathbf{P}\left(\left|\sum_{i=1}^n X_{ni} - \sum_{i=1}^n \bar{X}_{ni}\right| \geq \delta\right) &\leq \mathbf{P}\left(\sum_{i=1}^n X_{ni} \neq \sum_{i=1}^n \bar{X}_{ni}\right) \\ &\leq \mathbf{P}\left(\bigcup_{i=1}^n \{X_{ni} \neq \bar{X}_{ni}\}\right) \leq \sum_{i=1}^n \mathbf{P}(X_{ni} \neq \bar{X}_{ni}) = \sum_{i=1}^n \mathbf{P}(|X_{ni}| \geq 1) \\ &= \sum_{i=1}^n \mathbf{P}(|X_i| \geq n) = n \mathbf{P}(|X_1| \geq n) \rightarrow 0 \end{aligned}$$

por la condición (5.10). Queda demostrado entonces (5.12). \square

Veamos ahora un corolario del teorema 5.5, de sencilla formulación.

Teorema 5.6. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes dos a dos e idénticamente distribuidas. Supongamos que existe la esperanza matemática $\mathbf{E} X_1$. Entonces*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E} X_1.$$

Demostración. Demostremos primero el teorema 5.6 en el caso particular en el que $\mathbf{E} X_1 = 0$, verificando las condiciones en el teorema 5.5. Sea $V(x)$ la función de distribución de la variable aleatoria X_1 . La condición (5.11) se cumple en forma evidente. Para verificar la condición (5.10), veamos que

$$\mathbf{P}(|X_1| \geq n) = \int_{\{|x| \geq n\}} dV(x) \leq \frac{1}{n} \int_{\{|x| \geq n\}} |x| dV(x),$$

por ésto,

$$n \mathbf{P}(|X_1| \geq n) \leq \int_{\{|x| \geq n\}} |x| dV(x) \rightarrow 0$$

cuando $n \rightarrow \infty$, porque $\int_{-\infty}^{\infty} |x| dV(x) = \mathbf{E}|X_1| < \infty$. De esta forma se verifican todas las hipótesis del teorema 5.5, concluyendo la demostración en este caso particular.

Supongamos ahora que $\mathbf{E}X_1 = a \neq 0$. Consideremos la sucesión de variables aleatorias $\{Y_n\}$, donde $Y_n = X_n - a$ ($n = 1, 2, \dots$). Como $\{X_n\}$ es una sucesión de variables aleatorias independientes dos a dos, también lo es la sucesión $\{Y_n\}$. Además $\mathbf{E}Y_n = \mathbf{E}X_n - a = 0$. Según demostramos en la primera parte, tiene lugar la convergencia, $\sum_{i=1}^n Y_i/n \xrightarrow{\mathbf{P}} 0$, que claramente equivale a $\sum_{i=1}^n X_i/n \xrightarrow{\mathbf{P}} a$, lo que concluye la demostración. \square

El teorema 5.6 fue obtenido por A. Ya. Jínchin, bajo la hipótesis más restrictiva de independencia mutua, en vez de independencia dos a dos, de las variables aleatorias X_1, \dots, X_n .

En el caso en que las variables aleatorias sean idénticamente distribuidas, el teorema de Chebishev conduce a un resultado más débil que el teorema 5.6, dado que en el teorema de Chebishev se exige la existencia de varianzas, y no únicamente la existencia de esperanzas matemáticas.

5.3. Ejercicios

1. Sean $\{X_n\}$ e $\{Y_n\}$ dos sucesiones de variables aleatorias, tales que $X_n \xrightarrow{\mathbf{P}} 0$, $Y_n \xrightarrow{\mathbf{P}} 0$. Demostrar que $X_n + Y_n \xrightarrow{\mathbf{P}} 0$.
2. Sea X, X_1, X_2, \dots una sucesión de variables aleatorias que converge en probabilidad a una variable aleatoria X . Sean $\{a_n\}$ y $\{b_n\}$ sucesiones numéricas, tales que $a_n \rightarrow a > 0$, $b_n \rightarrow b$. Demostrar que $a_n X_n + b_n \xrightarrow{\mathbf{P}} aX + b$.
3. Sea $\{X_n\}$ una sucesión de variables aleatorias; Y una variable aleatoria tal que $\mathbf{P}(Y = 0) = 1$. Demostrar que $X_n \xrightarrow{\mathbf{P}} 0$ si y solo si $X_n \xrightarrow{d} Y$.
4. Construir un ejemplo, demostrando que la convergencia $X_n \xrightarrow{d} X$ no implica la convergencia $X_n \xrightarrow{\mathbf{P}} X$.
5. Sean $\{X_n\}$ e $\{Y_n\}$ dos sucesiones de variables aleatorias, tales que $X_n \xrightarrow{\mathbf{P}} X$, $Y_n \xrightarrow{\mathbf{P}} Y$. Demostrar que $X_n + Y_n \xrightarrow{\mathbf{P}} X + Y$.

6. Sean $\{X_n\}$ e $\{Y_n\}$ dos sucesiones de variables aleatorias, tales que $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{\mathbf{P}} 0$. Demostrar que $X_n + Y_n \xrightarrow{d} X$.

7. Sean $\{X_n\}$ e $\{Y_n\}$ dos sucesiones de variables aleatorias, tales que $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{\mathbf{P}} a$, donde X es una variable aleatoria, a una constante. Demostrar la siguiente proposición (*teorema de Cramér-Slutsky*): Si se cumplen las condiciones anteriores, tienen lugar las afirmaciones: $X_n + Y_n \xrightarrow{d} X + a$, $X_n Y_n \xrightarrow{d} aX$, $X_n/Y_n \xrightarrow{d} X/a$ si $a \neq 0$.

8. Sea $\{X_n\}$ una sucesión de variables aleatorias, tales que $X_n \xrightarrow{\mathbf{P}} a$, donde a es una constante. Sea $g(x)$ una función, continua en el punto a . Demostrar que $g(X_n) \xrightarrow{\mathbf{P}} g(a)$.

9. Sean X, X_1, X_2, \dots una sucesión de variables aleatorias, y $g(x)$ una función definida y continua en la recta real. Demostrar que si $X_n \xrightarrow{\mathbf{P}} X$, entonces $g(X_n) \xrightarrow{\mathbf{P}} g(X)$.

10. Sea X, X_1, X_2, \dots una sucesión de variables aleatorias, tal que $X_n \rightarrow X$ en media cuadrática. Demostrar que $\mathbf{E} X_n \rightarrow \mathbf{E} X$, y que $\mathbf{E} X_n^2 \rightarrow \mathbf{E} X^2$.

11. Sean $\{X_n\}$, $\{Y_n\}$ sucesiones de variables aleatorias, y sea $g(x, y)$ una función definida y continua en \mathbb{R}^2 . Demostrar que si $X_n \xrightarrow{\mathbf{P}} X$, $Y_n \xrightarrow{\mathbf{P}} Y$, entonces $g(X_n, Y_n) \xrightarrow{\mathbf{P}} g(X, Y)$.

12. Sea $\{F_n(x)\}$ una sucesión de funciones de distribución, y sea $F(x)$ una función de distribución continua. Demostrar que si $F_n(x) \rightarrow F(x)$ para todo x , esta convergencia es uniforme en toda la recta real (*teorema de Pólya*). Sugerencia: Dado $\varepsilon > 0$ arbitrario, elegir reales $x_0 < x_1 < \dots < x_N$, en forma conveniente, y analizar la convergencia en cada uno de los $N + 2$ intervalos que estos números determinan.

13. Sea $\{X_n\}$ una sucesión de variables aleatorias, independientes dos a dos, tales que

$$\mathbf{P}(X_n = 2^n) = \mathbf{P}(X_n = -2^n) = 2^{-2n-1}, \quad \mathbf{P}(X_n = 0) = 1 - 2^{-2n},$$

para cada n . ¿Es aplicable la ley de los grandes números a esta sucesión?

14. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes dos a dos, tales que

$$\mathbf{P}(X_n = n^{1/4}) = \mathbf{P}(X_n = -n^{1/4}) = \mathbf{P}(X_n = 0) = 1/3,$$

para cada n . Demostrar que para esta sucesión es aplicable la ley de los grandes números.

15. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que

$$\mathbf{P}(X_n = n^\gamma) = \mathbf{P}(X_n = -n^\gamma) = 1/2,$$

para cada n , donde $\gamma < 1/2$. ¿Es aplicable la ley de los grandes números a esta sucesión?

16. Sea $\{X_n\}$ una sucesión de variables aleatorias, tal que $\mathbf{E} X_n = 0$, $\mathbf{E} |X_n| = 1/n$ para cada n . Demostrar que para esta sucesión es aplicable la ley de los grandes números.

17. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que

$$\mathbf{P}(X_n = n) = \mathbf{P}(X_n = -n) = 1/2$$

para cada n . ¿Es aplicable la ley de los grandes números a esta sucesión?

Capítulo 6

Funciones características

Una función característica es una función que toma valores complejos y tiene argumento real. Definida a partir de una variable aleatoria X , *caracteriza* la distribución $F(x)$ de esta variable aleatoria. Las funciones características son especialmente adecuadas para el estudio de la convergencia débil de variables aleatorias independientes, y serán utilizadas a lo largo del capítulo 7. Dadas dos variables aleatorias X e Y , definimos la *variable aleatoria compleja* $Z = X + iY$, donde $i = \sqrt{-1}$ es la unidad imaginaria. Si las variables aleatorias X, Y tienen esperanzas respectivas $\mathbf{E}X, \mathbf{E}Y$, definimos la esperanza matemática de la variable aleatoria compleja Z mediante $\mathbf{E}Z = \mathbf{E}X + i\mathbf{E}Y$. No es difícil verificar (tomando partes real e imaginaria), que si a, b son dos números complejos, se tiene $\mathbf{E}(aZ + b) = a\mathbf{E}Z + b$; y que si Z_1, Z_2 son dos variables aleatorias complejas, se tiene $\mathbf{E}(Z_1 + Z_2) = \mathbf{E}Z_1 + \mathbf{E}Z_2$. Si $z = a + ib$ es un número complejo, designamos $\bar{z} = a - ib$ el *complejo conjugado* de z .

6.1. Definiciones y primeras propiedades

Consideremos una variable aleatoria X definida en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. Llamamos *función característica* de la variable aleatoria X a la función $f(t)$, definida para todo t real, mediante la igualdad

$$f(t) = \mathbf{E} e^{itX}. \quad (6.1)$$

La fórmula (6.1) es equivalente a

$$f(t) = \mathbf{E} \cos tX + i \mathbf{E} \operatorname{sen} tX. \quad (6.2)$$

Como las variables aleatorias $\cos tX$ y $\sin tX$ están acotadas para todo t real, sus esperanzas matemáticas existen. Por ésto, la función característica de una variable aleatoria arbitraria X está correctamente definida para todo t real.

De acuerdo a la definición de esperanza matemática, tenemos

$$f(t) = \int_{\Omega} e^{itX} d\mathbf{P}. \quad (6.3)$$

Si la variable aleatoria X tiene función de distribución $F(x)$, aplicando la identidad (4.3) obtenemos

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x). \quad (6.4)$$

Consideremos ahora los dos tipos más importantes de distribuciones. Si la variable aleatoria X tiene distribución discreta, y toma los valores x_1, x_2, \dots con probabilidades p_1, p_2, \dots respectivamente, entonces

$$f(t) = \sum_k e^{itx_k} p_k, \quad (6.5)$$

como se obtiene de aplicar cualquiera de las fórmulas (6.1), (6.2), ó (6.3).

Si X tiene distribución absolutamente continua, con densidad dada por $p(x)$, entonces

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx, \quad (6.6)$$

como se obtiene de aplicar (6.4).

Calculemos las funciones características de variables aleatorias con distribuciones de ambos tipos, en los casos mas importantes.

Ejemplo 6.1. Supongamos que la variable aleatoria X tiene distribución degenerada, es decir, existe una constante c tal que $\mathbf{P}(X = c) = 1$. Aplicando la fórmula (6.5), tenemos

$$f(t) = \mathbf{E} e^{itX} = e^{itc}.$$

En particular, si $\mathbf{P}(X = 0) = 1$, tenemos $f(t) = 1$ para todo t real.

Ejemplo 6.2. Sea X una variable aleatoria con distribución binomial con parámetros (n, p) . Aplicando (6.5), obtenemos

$$f(t) = \sum_{m=0}^n e^{itm} \binom{n}{m} p^m q^{n-m} = \sum_{m=0}^n \binom{n}{m} (pe^{it})^m q^{n-m} = (pe^{it} + q)^n,$$

donde $q = 1 - p$.

Ejemplo 6.3. Si X tiene distribución de Poisson con parámetro $\lambda > 0$, aplicando la fórmula (6.5), obtenemos

$$f(t) = \mathbf{E} e^{itX} = \sum_{m=0}^{\infty} e^{itm} \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(e^{it}\lambda)^m}{m!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}.$$

Ejemplo 6.4. Consideremos una variable aleatoria X con distribución normal estándar, con densidad dada por $p(x) = e^{-x^2/2}/\sqrt{2\pi}$. Aplicando la fórmula (6.6), tenemos

$$f(t) = \mathbf{E} e^{itX} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx-x^2/2} dx.$$

Para calcular la integral anterior derivamos con respecto de t , y obtenemos

$$f'(t) = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{itx-x^2/2} dx = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} d(-e^{-x^2/2}).$$

Luego de integrar por partes, resulta

$$f'(t) = \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{itx-x^2/2} dx = -t f(t).$$

En consecuencia, $(\ln f(t))' = -t$, $\ln f(t) = -t^2/2 + C$. Como $f(0) = 1$, obtenemos que $C = 0$, y en conclusión

$$f(t) = e^{-t^2/2}.$$

Si X es una variable aleatoria con distribución normal con parámetros (a, σ) , entonces, como veremos en el ejemplo 6.6, su función característica está dada por

$$f(t) = e^{iat-\sigma^2 t^2/2}.$$

Ejemplo 6.5. Sea X una variable aleatoria con distribución exponencial con parámetro $\alpha > 0$. La densidad de esta variable aleatoria está dada por $p(x) = \alpha e^{-\alpha x}$ si $x \geq 0$, $p(x) = 0$ si $x < 0$. Por ésto, aplicando la fórmula (6.6), tenemos

$$f(t) = \mathbf{E} e^{itX} = \alpha \int_0^{\infty} e^{(it-\alpha)x} dx = \frac{\alpha}{\alpha - it}.$$

Estudiemos ahora las propiedades que verifica la función característica $f(t)$ de una variable aleatoria X con función de distribución $F(x)$. Las dos primeras propiedades son evidentes.

Propiedad 1. *Se tiene $f(0) = 1$.*

Propiedad 2. *Se tiene $|f(t)| \leq 1$ para cualquier $t \in \mathbb{R}$.*

Propiedad 3. *La función característica $f(t)$ es uniformemente continua en la recta real.*

Demostración. La demostración, basada en la fórmula (6.3), se adapta sin dificultad si tomamos como definición cualquiera de las fórmulas (6.4), (6.5) ó (6.6).

Sean t, h reales arbitrarios. De (6.3), tenemos

$$f(t+h) - f(t) = \int_{\Omega} e^{itX} (e^{ihX} - 1) d\mathbf{P}. \quad (6.7)$$

Utilizamos la acotación $|e^{iu} - 1| \leq |u|$, válida para todo u real, porque

$$|e^{iu} - 1| = \left| \int_0^u e^{ix} dx \right| \leq \int_0^{|u|} |e^{ix}| dx = |u|.$$

Sea $\varepsilon > 0$, arbitrario. Existe un real A que verifica: A y $-A$ son puntos de continuidad de $F(x)$; $1 - F(A) < \varepsilon/8$; $F(-A) < \varepsilon/8$. Tomando valor absoluto a ambos lados en (6.7), y designando $\mathbf{B} = \{\omega \in \Omega: |X(\omega)| \geq A\}$, obtenemos

$$\begin{aligned} |f(t+h) - f(t)| &\leq \int_{\Omega} |e^{ihX} - 1| d\mathbf{P} \leq 2 \int_{\mathbf{B}} d\mathbf{P} + \int_{\mathbf{B}^c} |e^{ihX} - 1| d\mathbf{P}, \\ &\leq 2\mathbf{P}(\mathbf{B}) + \int_{\mathbf{B}} |hX| d\mathbf{P} \leq 2\mathbf{P}(|X| \geq A) + A|h| \\ &\leq 2(1 - F(A) + F(-A)) + Ah \leq \frac{\varepsilon}{2} + Ah < \varepsilon, \end{aligned}$$

si tomamos $h < \varepsilon/(2A)$. Como la acotación es independiente de t , esto concluye la demostración. \square

Propiedad 4. *Consideremos la variable aleatoria $Y = aX + b$, donde a, b son constantes. Entonces, la función característica $g(t)$ de la variable aleatoria Y verifica $g(t) = e^{ibt} f(at)$, donde $f(t)$ es la función característica de la variable aleatoria X .*

Demostración. Utilizando la propiedad $e^{i(\alpha+\beta)} = e^{i\alpha}e^{i\beta}$ y aplicando la definición (6.2), tenemos

$$g(t) = \mathbf{E} e^{itY} = \mathbf{E} e^{it(aX+b)} = \mathbf{E}(e^{ibt}e^{iatX}) = e^{ibt} \mathbf{E} e^{iatX} = e^{ibt} f(at),$$

lo que concluye la demostración. \square

Ejemplo 6.6. Calculemos la función característica $g(t)$ de una variable aleatoria Y con función de distribución normal con parámetros (a, σ) .

Es claro que la variable aleatoria $X = (Y - a)/\sigma$ tiene distribución normal con parámetros $(0, 1)$, y por lo tanto función característica $f(t) = e^{-t^2/2}$, como vimos en el ejemplo 6.4. Aplicando la propiedad anterior, obtenemos

$$g(t) = e^{iat} f(\sigma t) = e^{iat - \sigma^2 t^2/2}.$$

Propiedad 5. Consideremos una variable aleatoria X para la cual existe $\alpha_k = \mathbf{E}(X^k)$, el momento de orden k , para algun natural $k \geq 1$. Entonces, su función característica $f(t)$ tiene derivadas continuas, para todo t real, hasta el orden k inclusive. Además

$$f^{(m)}(0) = i^m \alpha_m \quad (1 \leq m \leq k), \quad (6.8)$$

donde $\alpha_m = \mathbf{E}(X^m)$.

Demostración. Derivando formalmente m veces, con respecto de t , bajo el signo de integración en la fórmula (6.4), obtenemos

$$f^{(m)}(t) = \int_{-\infty}^{\infty} (ix)^m e^{itx} dF(x). \quad (6.9)$$

Es claro que como existe el momento de orden $m \leq k$ (proposición 4.1), tenemos

$$\left| \int_{-\infty}^{\infty} (ix)^m e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |x|^m dF(x) < \infty.$$

Esto permite demostrar que la derivación bajo el signo de integral es válida, y obtener la fórmula (6.9). Sustituyendo $t = 0$ en (6.9) se obtiene (6.8). \square

Dada una variable aleatoria X , si para algún natural $k \geq 1$ existe $\alpha_k = \mathbf{E}(X^k)$, el momento de orden k de la variable aleatoria, aplicando la

propiedad 5, el desarrollo de Taylor, y la igualdad $f(0) = 1$, se obtiene, que

$$f(t) = 1 + \sum_{m=1}^k \frac{\alpha_m}{m!} (it)^m + o(|t|^k) \quad (t \rightarrow 0). \quad (6.10)$$

En la demostración de la próxima propiedad, utilizamos el resultado siguiente.

Lema 6.1. *Consideremos dos variables aleatorias independientes X, Y , dos funciones reales $u(x), v(x)$, definidas en la recta real, y supongamos que existen $\mathbf{E}u(X)$ y $\mathbf{E}v(Y)$. Entonces, tiene lugar la identidad*

$$\mathbf{E}(u(X)v(Y)) = \mathbf{E}u(X) \mathbf{E}v(Y).$$

Demostración. Veremos la demostración en el caso en que ambas variables tienen distribución discreta, y en el caso en que tienen distribución absolutamente continua.

Supongamos primero que X toma los valores x_1, x_2, \dots , con probabilidades p_1, p_2, \dots , respectivamente; Y toma los valores y_1, y_2, \dots , con probabilidades q_1, q_2, \dots , respectivamente. Aplicando la proposición 3.1, obtenemos que $\mathbf{P}(X = x_k, Y = y_j) = \mathbf{P}(X = x_k) \mathbf{P}(Y = y_j) = p_k q_j$ ($k, j = 1, 2, \dots$). Por ésto, tenemos

$$\begin{aligned} \mathbf{E}(u(X)v(Y)) &= \sum_{k,j} u(x_k)v(y_j) \mathbf{P}(X = x_k, Y = y_j) \\ &= \sum_k u(x_k)p_k \sum_j v(y_j)q_j = \mathbf{E}u(X) \mathbf{E}v(Y). \end{aligned}$$

Si X e Y tienen distribución absolutamente continua y $r(x, y)$ designa la densidad del vector (X, Y) , aplicando la proposición 3.3, obtenemos que $r(x, y) = p(x)q(y)$, donde $p(x)$ es la densidad de la variable aleatoria X , $q(y)$ la densidad de la variable aleatoria Y . Por ésto, tenemos

$$\begin{aligned} \mathbf{E}(u(X)v(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x)v(y)r(x, y)dx dy \\ &= \int_{-\infty}^{\infty} u(x)p(x)dx \int_{-\infty}^{\infty} v(y)q(y)dy = \mathbf{E}u(X) \mathbf{E}v(Y), \end{aligned}$$

concluyendo la demostración. □

Observación. El lema recién formulado es válido también en el caso en que las funciones de $u(x)$ y $v(x)$ de argumento real, tomen valores complejos, es decir, si tenemos

$$u(x) = u_1(x) + iu_2(x), \quad v(x) = v_1(x) + iv_2(x),$$

donde $u_k(x)$ y $v_k(x)$ son funciones de argumento real, que toman valores reales ($k = 1, 2$). Esto es sencillo de demostrar, aplicando el lema anterior a las partes real e imaginaria del producto $u(X)v(Y)$.

Propiedad 6. *Consideremos dos variables aleatorias independientes X e Y , con funciones características $f(t)$ y $g(t)$ respectivamente. Sea $h(t)$ la función característica de la suma $X + Y$. Entonces, se verifica $h(t) = f(t)g(t)$.*

Demostración. Tenemos

$$h(t) = \mathbf{E} e^{it(X+Y)} = \mathbf{E} \left(e^{itX} e^{itY} \right) = \mathbf{E} e^{itX} \mathbf{E} e^{itY} = f(t)g(t),$$

en vista de la observación posterior al lema 6.1. □

Es válida la siguiente generalización de la propiedad recién demostrada: si X_1, X_2, \dots, X_n es un conjunto de variables aleatorias mutuamente independientes, con funciones características $f_1(t), f_2(t), \dots, f_n(t)$ respectivamente, entonces, la función característica $h(t)$ de la suma $X_1 + X_2 + \dots + X_n$ es igual al producto de las funciones características de los sumandos:

$$h(t) = f_1(t)f_2(t) \cdots f_n(t).$$

Propiedad 7. *Para todo t real, se verifica $f(-t) = \overline{f(t)}$.*

La propiedad anterior se obtiene de la igualdad

$$f(-t) = \mathbf{E} e^{-itX} = \mathbf{E} \overline{e^{itX}} = \overline{\mathbf{E} e^{itX}} = \overline{f(t)}.$$

Definición 6.1. *Una variable aleatoria X y su distribución $F(x)$ se dicen simétricas cuando las funciones de distribución de las variables aleatorias X y $-X$ son idénticas.*

Propiedad 8. *Si la variable aleatoria X es simétrica, su función característica $f(t)$ es una función real.*

A la conclusión de la propiedad anterior conducen las igualdades

$$f(t) = \mathbf{E} e^{itX} = \mathbf{E} e^{it(-X)} = f(-t) = \overline{f(t)}.$$

En la sección 2 demostraremos el recíproco de la propiedad anterior: si la función característica de una variable aleatoria X es real, la variable aleatoria X es simétrica.

6.2. Fórmula de inversión. Teorema de unicidad

Teorema 6.1. *Consideremos una variable aleatoria X con función de distribución $F(x)$, y función característica $f(t)$. Sean x_1, x_2 dos puntos de continuidad de $F(x)$. Entonces, tiene lugar la igualdad*

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t) dt. \quad (6.11)$$

La igualdad (6.11) se denomina fórmula de inversión.

Demostración. Comenzamos introduciendo la función auxiliar

$$R(h, T) = \int_0^T \frac{\text{sen } ht}{t} dt = \int_0^{hT} \frac{\text{sen } u}{u} du.$$

Del cálculo integral, son conocidas las siguientes afirmaciones:

$$\int_0^\infty \frac{\text{sen } u}{u} du = \frac{\pi}{2}, \quad \left| \int_0^x \frac{\text{sen } u}{u} du \right| \leq C \quad (x \geq 0),$$

de donde obtenemos, que

$$\lim_{T \rightarrow \infty} R(h, T) = \begin{cases} -\pi/2, & \text{si } h < 0, \\ \pi/2, & \text{si } h > 0. \end{cases} \quad (6.12)$$

Es importante observar, que esta convergencia es uniforme en los intervalos de la forma $(-\infty, \delta]$, y $[\delta, \infty)$, para todo $\delta > 0$.

La segunda etapa de la demostración, consiste en representar a la integral en (6.11), que designamos I , en términos de la función $R(h, T)$.

Aplicando la definición (6.4), e intercambiando el orden de integración (dado que el integrando es una función continua y acotada), tenemos

$$\begin{aligned}
 I &= \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} f(t) dt = \int_{-T}^T \frac{e^{-itx_2} - e^{-itx_1}}{-it} \left(\int_{-\infty}^{\infty} e^{ity} dF(y) \right) dt \\
 &= \int_{-\infty}^{\infty} \left(\int_{-T}^T \frac{e^{it(y-x_2)} - e^{it(y-x_1)}}{-it} dt \right) dF(y) \\
 &= \int_{-\infty}^{\infty} \left(\int_{-T}^T \frac{\operatorname{sen}(t(y-x_2)) - \operatorname{sen}(t(y-x_1))}{-t} dt \right) dF(y) \\
 &= 2 \int_{-\infty}^{\infty} (R(y-x_1, T) - R(y-x_2, T)) dF(y).
 \end{aligned}$$

donde utilizamos que $\int_{-T}^T (\cos(\alpha t)/t) dt = 0$ para todo α real, para obtener la última igualdad. Respecto del comportamiento asintótico del último integrando, tomando por ejemplo $x_1 < x_2$, en vista de (6.12), tenemos

$$\lim_{T \rightarrow \infty} R(y-x_1, T) - R(y-x_2, T) = \begin{cases} 0, & \text{si } y < x_1, \\ \pi, & \text{si } x_1 < y < x_2, \\ 0, & \text{si } x_2 < y. \end{cases} \quad (6.13)$$

La última etapa de la demostración consiste en verificar que el límite de I cuando $T \rightarrow \infty$ es la integral del límite obtenido en (6.13). Para ésto elegimos $\delta > 0$, de forma que $x_1 + \delta < x_2 - \delta$, y consideramos la integral I como la suma de cinco integrales, designadas I_k ($i = 1, \dots, 5$), en los intervalos de integración $(-\infty, x_1 - \delta]$, $(x_1 - \delta, x_1 + \delta]$, $(x_1 + \delta, x_2 - \delta]$, $(x_2 - \delta, x_2 + \delta]$, y $(x_2 + \delta, \infty)$. Tenemos entonces $I = \sum_{i=1}^5 I_k$.

Sea $\varepsilon > 0$ arbitrario. Tenemos

$$I_1 = 2 \int_{(-\infty, x_1 - \delta]} (R(y-x_2, T) - R(y-x_1, T)) dF(y).$$

Como, en vista de (6.13), el integrando converge uniformemente a cero, obtenemos que $|I_1| < \varepsilon$ si T es suficientemente grande. Una situación análoga ocurre con I_5 , por lo que, $|I_5| < \varepsilon$ si T es suficientemente grande.

Para la segunda integral, tenemos

$$\begin{aligned}
 |I_2| &\leq 2 \int_{(x_1 - \delta, x_1 + \delta]} (R(y-x_2, T) - R(y-x_1, T)) dF(y) \\
 &\leq 4C(F(x_1 + \delta) - F(x_1 - \delta)) < \varepsilon,
 \end{aligned}$$

si δ es suficientemente pequeño (independientemente de T), dado que x_1 es un punto de continuidad de $F(x)$. La situación con I_4 es análoga, y por ésto $|I_4| < \varepsilon$ si δ es suficientemente pequeño. Por último, como la convergencia en (6.13) es uniforme, para la tercer integral tenemos

$$\begin{aligned} I_3 &= 2 \int_{(x_1+\delta, x_2-\delta]} (R(y-x_2, T) - R(y-x_1, T)) dF(y) \\ &\rightarrow 2\pi(F(x_2 - \delta) - F(x_1 + \delta)), \end{aligned}$$

si $T \rightarrow \infty$. En conclusión, para todo δ suficientemente pequeño, tenemos

$$\begin{aligned} \left| I - 2\pi(F(x_2 - \delta) - F(x_1 + \delta)) \right| &\leq |I_1| + |I_2| + |I_4| + |I_5| \\ &\quad + \left| I - 2\pi(F(x_2 - \delta) - F(x_1 + \delta)) \right| < 5\varepsilon, \end{aligned}$$

si T es suficientemente grande. Como x_1 y x_2 son puntos de continuidad de la función $F(x)$, esto concluye la demostración. \square

Teorema 6.2 (Unicidad). *Sean $f(t), g(t)$ las funciones características correspondientes a dos funciones de distribución $F(x), G(x)$. Supongamos que $f(t) = g(t)$ para todo t real. Entonces, se verifica $F(x) = G(x)$ para todo x real.*

Demostración. Sea \mathbf{C} el conjunto de los puntos en el que ambas funciones $F(x)$ y $G(x)$ son continuas. Como $F(x)$ y $G(x)$ son funciones de distribución, el complemento del conjunto \mathbf{C} es, a lo sumo, numerable. Sean entonces x, y_1, y_2, \dots puntos de \mathbf{C} , tales que $y_n \rightarrow -\infty$ ($n \rightarrow \infty$). Aplicando el teorema 6.1, obtenemos que $F(x) - F(y_n) = G(x) - G(y_n)$, y tomando límite si $n \rightarrow \infty$ en la igualdad anterior, resulta

$$F(x) = G(x) \text{ para todo } x \text{ en } \mathbf{C}. \quad (6.14)$$

Sea ahora z un real arbitrario. Consideremos $x_1 > x_2 > \dots$ puntos de \mathbf{C} , tales que $x_n \rightarrow z$ ($n \rightarrow \infty$). En vista de (6.14), tenemos $F(x_n) = G(x_n)$. Como ambas funciones de distribución son continuas por la derecha, al tomar límite si $n \rightarrow \infty$ en la igualdad anterior, obtenemos la igualdad $F(z) = G(z)$. Como z es arbitrario, esto concluye la demostración. \square

De acuerdo al teorema 6.2, denominado *teorema de unicidad*, la función característica de una variable aleatoria define unívocamente (es decir “caracteriza”) su función de distribución. Veamos algunas aplicaciones de este teorema.

Ejemplo 6.7. Consideremos dos variables aleatorias independientes X_1 y X_2 , cada una de las cuales tiene distribución normal con parámetros (a_1, σ_1) y (a_2, σ_2) respectivamente. Con la ayuda del teorema de unicidad es sencillo demostrar que la suma $X_1 + X_2$ tiene distribución normal (como vimos en la sección 3.4).

La función característica de X_k es $f_k(t) = e^{ia_k t - \sigma_k^2 t^2 / 2}$ ($k = 1, 2$). Como las variables aleatorias son independientes, la función característica de la suma $X_1 + X_2$ es

$$\begin{aligned} g(t) &= \mathbf{E} e^{it(X_1+X_2)} = \mathbf{E} e^{itX_1} \mathbf{E} e^{itX_2} \\ &= f_1(t)f_2(t) = e^{i(a_1+a_2)t - (\sigma_1^2 + \sigma_2^2)t^2/2}. \end{aligned}$$

Es claro que la función $g(t)$ coincide con la función característica de una variable aleatoria con distribución normal, con parámetros $(a_1 + a_2, (\sigma_1^2 + \sigma_2^2)^{1/2})$. Aplicando el teorema 6.2 se deduce que la suma $X_1 + X_2$ es una variable aleatoria con distribución normal con éstos parámetros.

Ejemplo 6.8. Consideremos dos variables aleatorias independientes X e Y , cada una de las cuales tiene distribución exponencial con parámetros $\alpha > 0$ y $\beta > 0$ respectivamente. Veamos que la variable aleatoria $Z = X - Y$ tiene densidad, dada por

$$p(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha x} & \text{si } x > 0, \\ \frac{\alpha\beta}{\alpha+\beta} e^{\beta x} & \text{si } x \leq 0. \end{cases}$$

Por una parte, utilizando el ejemplo 6.5, tenemos

$$\mathbf{E} e^{itZ} = \mathbf{E} e^{itX} \mathbf{E} e^{-itY} = \frac{\alpha}{\alpha - it} \frac{\beta}{\beta + it}. \quad (6.15)$$

Por otra parte,

$$\begin{aligned} \int_{-\infty}^{\infty} e^{itx} p(x) dx &= \frac{\alpha\beta}{\alpha + \beta} \left(\int_0^{\infty} e^{(it-\alpha)x} dx + \int_{-\infty}^0 e^{(it+\beta)x} dx \right) \\ &= \frac{\alpha\beta}{\alpha + \beta} \left(\frac{1}{\alpha - it} + \frac{1}{\beta + it} \right) = \frac{\alpha}{\alpha - it} \frac{\beta}{\beta + it}. \end{aligned} \quad (6.16)$$

Como los resultados en (6.15) y (6.16) coinciden, del teorema de unicidad se deduce que $p(x)$ es la densidad de Z .

Veamos un corolario más del teorema de unicidad: si la función característica $f(t)$ de una variable aleatoria dada es real, entonces la variable aleatoria es simétrica. En efecto, como $f(t)$ es real, aplicando la propiedad 7, tenemos

$$f(t) = \overline{f(t)} = f(-t).$$

Por otra parte, $f(-t) = \mathbf{E} e^{-itX}$ es la función característica de la variable aleatoria $-X$ en el punto t . De la coincidencia de las funciones características se obtiene la igualdad de las distribuciones de las variables aleatorias X y $-X$, aplicando el teorema 6.2. Concluimos que X es simétrica.

En vista de lo recién demostrado y de la propiedad 8, llegamos a la siguiente conclusión: una variable aleatoria es simétrica si y solo si su función característica es real.

6.3. Teoremas de Helly

Los teoremas de Helly juegan un importante rol en la demostración de los teoremas de convergencia de funciones características, que estudiamos en la sección 4. Dado que no se incluyen en los cursos habituales de cálculo, los presentamos aquí con sus correspondientes demostraciones.

Definición 6.2. *Consideremos funciones $F(x), F_1(x), F_2(x), \dots$, acotadas y no decrecientes.*

(a) *Decimos que la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$ y escribimos $F_n \rightarrow F$, si $F_n(x) \rightarrow F(x)$ ($n \rightarrow \infty$) para todo punto x de continuidad de la función $F(x)$.*

(b) *Decimos que la sucesión $\{F_n(x)\}$ converge completamente a $F(x)$ y escribimos $F_n \Rightarrow F$, si $\{F_n(x)\}$ converge débilmente a $F(x)$ (es decir, si $F_n \rightarrow F$) y además¹ $F_n(-\infty) \rightarrow F(-\infty)$, $F_n(\infty) \rightarrow F(\infty)$ si $n \rightarrow \infty$.*

Observemos que si $F(x), F_1(x), F_2(x), \dots$ son funciones de distribución, la convergencia débil definida en (a) coincide con la convergencia débil de variables aleatorias definida en la sección 5.1. Además, en este caso, las definiciones (a) y (b) coinciden. Sin embargo, en el caso general esta equivalencia no es cierta, como se ve en el siguiente ejemplo.

¹Designamos $G(-\infty) = \lim_{x \rightarrow -\infty} G(x)$, cuando existe este límite para una cierta función $G(x)$. Análogamente, designamos $G(\infty) = \lim_{x \rightarrow \infty} G(x)$.

Consideremos, para cada $n = 1, 2, \dots$, la función

$$F_n(x) = \begin{cases} 0, & \text{si } x \leq -n, \\ 1/2, & \text{si } -n < x \leq n, \\ 1, & \text{si } n \leq x. \end{cases}$$

Es claro que $F_n(x) \rightarrow 1/2$ ($n \rightarrow \infty$) para todo x real, y en consecuencia la sucesión $\{F_n(x)\}$ considerada converge débilmente a la función $F(x) = 1/2$. Sin embargo, la convergencia completa no tiene lugar, dado que $F_n(-\infty) = 0$, $F_n(+\infty) = 1$ para todo n , y tenemos $F(-\infty) = F(+\infty) = 1/2$.

Teorema 6.3 (Helly). *Consideremos una sucesión $\{F_n(x)\}$ de funciones no decrecientes. Supongamos que existen constantes A y B tales que se verifica $A \leq F_n(x) \leq B$ para todo x real y para todo n . Entonces, la sucesión dada contiene una subsucesión $\{F_{n_k}(x)\}$ que converge débilmente a una cierta función $F(x)$, no decreciente y acotada.*

En la demostración de este teorema utilizamos el siguiente resultado.

Lema 6.2. *Si $F_n(x) \rightarrow F(x)$ para todo $x \in \mathbf{D}$, donde \mathbf{D} es un conjunto denso en la recta real, entonces $F_n \rightarrow F$.*

Demostración del lema. Sea x un punto de continuidad de $F(x)$. Como \mathbf{D} es denso, existen dos sucesiones $\{x'_k\}$ y $\{x''_k\}$ que verifican $x'_k < x < x''_k$ para todo k y $\lim_{k \rightarrow \infty} x'_k = \lim_{k \rightarrow \infty} x''_k = x$. Para cada k y cada n , tenemos

$$F_n(x'_k) \leq F_n(x) \leq F_n(x''_k).$$

Como $\lim_{n \rightarrow \infty} F_n(x'_k) = F(x'_k)$ y $\lim_{n \rightarrow \infty} F_n(x''_k) = F(x''_k)$, de las desigualdades anteriores, si $n \rightarrow \infty$, obtenemos

$$F(x'_k) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x''_k).$$

Hagamos ahora tender k a infinito. Dado que x es un punto de continuidad de $F(x)$, se verifica $\lim_{k \rightarrow \infty} F(x'_k) = \lim_{k \rightarrow \infty} F(x''_k) = F(x)$, por lo que

$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x).$$

Entonces, el $\lim_{n \rightarrow \infty} F_n(x)$ existe, y vale $F(x)$. Como el punto de continuidad es arbitrario, esto concluye la demostración. \square

Demostración del teorema 6.3 de Helly. Sea \mathbf{D} un conjunto denso numerable de números reales x'_1, x'_2, \dots . La sucesión numérica $\{F_n(x'_1)\}$ es acotada, por lo que contiene una subsucesión $\{F_{1n}(x'_1)\}$ que converge a un cierto límite, que designamos $F(x'_1)$.

La sucesión $\{F_{1n}(x'_2)\}$ también contiene una subsucesión $\{F_{2n}(x'_2)\}$, convergente a un cierto límite, que designamos $F(x'_2)$. Además, se verifica $\lim_{n \rightarrow \infty} F_{2n}(x'_1) = F(x'_1)$.

Continuando este proceso, obtenemos, que para cualquier natural k , existen k sucesiones $\{F_{kn}(x'_i)\}$ ($i = 1, \dots, k$) para las cuales se verifica $\lim_{n \rightarrow \infty} F_{kn}(x'_i) = F(x'_i)$ ($i = 1, \dots, k$).

Consideremos ahora la sucesión diagonal compuesta por las funciones $\{F_{nn}(x)\}$. Sea $x'_k \in \mathbf{D}$. Es claro que $\lim_{n \rightarrow \infty} F_{nn}(x'_k) = F(x'_k)$, dado que $\{F_{nn}(x'_k)\}$ es una subsucesión de la sucesión numérica $\{F_{kn}(x'_k)\}$, si $n \geq k$. Hemos así definido una función $F(x)$ en el conjunto \mathbf{D} . Si $x < y$ son dos puntos de \mathbf{D} , entonces $F(x) = \lim_{n \rightarrow \infty} F_{nn}(x) \leq \lim_{n \rightarrow \infty} F_{nn}(y) = F(y)$, y la función $F(x)$ es no decreciente en \mathbf{D} . Es claro también que $A \leq F(x) \leq B$. Estas propiedades permiten extender la función $F(x)$ a toda la recta real, conservando las propiedades mencionadas. Estamos entonces en condiciones de aplicar el lema 6.2, para concluir la demostración del teorema. \square

Observación. Se puede ver que la función límite $F(x)$ puede elegirse continua por la derecha, si definimos $F(x) = \lim_{n \rightarrow \infty} F(x_n)$, donde $\{x_n\} \in \mathbf{D}$, $x_n \rightarrow x$ ($n \rightarrow \infty$), y $x_n \geq x$ para todo n .

Teorema 6.4 (Helly). *Consideremos funciones no decrecientes y acotadas $F(x), F_1(x), F_2(x), \dots$ tales que $F_n \Rightarrow F$, y una función $g(x)$ continua y acotada. Entonces*

$$\int_{-\infty}^{\infty} g(x) dF_n(x) \rightarrow \int_{-\infty}^{\infty} g(x) dF(x) \quad (n \rightarrow \infty).$$

Demostración. Sea $\varepsilon > 0$ arbitrario. Designemos

$$G = \sup_{x \in \mathbb{R}} |g(x)|, \quad C = F(\infty) - F(-\infty).$$

Como $\lim_{x \rightarrow -\infty} F(x) = F(-\infty)$, $\lim_{x \rightarrow \infty} F(x) = F(\infty)$, existen $a < b$, puntos de continuidad de $F(x)$, tales que se verifica

$$F(\infty) - F(b) < \varepsilon / (3G), \quad F(a) - F(-\infty) < \varepsilon / (3G). \quad (6.17)$$

Como $g(x)$ es una función continua, existe una partición del intervalo $[a, b]$, designada $a = x_0 < x_1 < \dots < x_N = b$, formada por puntos de continuidad de $F(x)$, y tales que se verifica $|g(x) - g(x_k)| < \delta$ si $x \in (x_{k-1}, x_k)$ ($k = 1, \dots, N$).

Consideremos la función auxiliar

$$g_0(x) = \begin{cases} g(x_k), & \text{si } x \in (x_{k-1}, x_k] \text{ } (k = 1, \dots, N), \\ 0, & \text{en otro caso,} \end{cases}$$

que podemos también definir como $g_0(x) = \sum_{k=1}^N g(x_k) \mathbf{1}_{(x_{k-1}, x_k]}(x)$. Sean

$$I = \int_{-\infty}^{\infty} g(x) dF(x), \quad I_n = \int_{-\infty}^{\infty} g(x) dF_n(x).$$

Sumando y restando, obtenemos

$$I_n - I = \int_{-\infty}^{\infty} (g(x) - g_0(x)) dF_n(x) \quad (6.18)$$

$$+ \int_{-\infty}^{\infty} g_0(x) dF_n(x) - \int_{-\infty}^{\infty} g_0(x) dF(x) \quad (6.19)$$

$$+ \int_{-\infty}^{\infty} (g_0(x) - g(x)) dF(x) \quad (6.20)$$

$$= S_1 + S_2 + S_3.$$

Acotemos cada uno de los tres sumandos anteriores. Para S_3 en (6.20), tenemos

$$\begin{aligned} |S_3| &\leq \int_{-\infty}^a |g(x)| dF(x) + \int_a^b |g(x) - g_0(x)| dF(x) + \int_b^{\infty} |g(x)| dF(x) \\ &\leq G(F(a) - F(-\infty)) + \delta(F(b) - F(a)) + G(F(\infty) - F(b)) \quad (6.21) \\ &\leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon, \end{aligned}$$

en vista de (6.17) y la desigualdad $F(b) - F(a) \leq C$.

Para S_1 en (6.18), cambiando F_n por F , obtenemos

$$\begin{aligned} |S_1| &\leq G(F_n(a) - F_n(-\infty)) \\ &\quad + \delta(F_n(b) - F_n(a)) + G(F_n(\infty) - F_n(b)) < \varepsilon, \end{aligned} \quad (6.22)$$

si n es suficientemente grande, dado que, por la convergencia completa $F_n \Rightarrow F$, la cota obtenida en (6.22) converge a la cota en (6.21).

Finalmente, también utilizando la convergencia completa $F_n \Rightarrow F$, obtenemos $S_2 \rightarrow 0$ ($n \rightarrow \infty$) en (6.19), porque

$$\begin{aligned} \int_{-\infty}^{\infty} g_0(x) dF_n(x) &= \sum_{k=1}^N g(x_k) (F_n(x_k) - F_n(x_{k-1})) \\ &\rightarrow \sum_{k=1}^N g(x_k) (F(x_k) - F(x_{k-1})) = \int_{-\infty}^{\infty} g_0(x) dF(x), \end{aligned}$$

si $n \rightarrow \infty$. En conclusión, para n suficientemente grande, tenemos

$$\left| \int_{-\infty}^{\infty} g(x) dF_n(x) - \int_{-\infty}^{\infty} g(x) dF(x) \right| < \varepsilon + \varepsilon + \varepsilon = 3\varepsilon,$$

lo que concluye la demostración. \square

6.4. Relación entre la convergencia de distribuciones y de funciones características

El objetivo de esta sección es la demostración del siguiente resultado.

Teorema 6.5. *Consideremos las funciones de distribución*

$$F(x), F_1(x), F_2(x), \dots$$

con sus correspondientes funciones características

$$f(t), f_1(t), f_2(t), \dots$$

Entonces, la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$ (es decir $F_n \rightarrow F$), si y solo si se verifica

$$f_n(t) \rightarrow f(t) \quad (n \rightarrow \infty) \text{ para todo } t \text{ real.} \quad (6.23)$$

Demostración. Supongamos primero que $F_n \rightarrow F$. Tenemos

$$f_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x), \quad f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x).$$

Como $e^{itx} = \cos tx + i \operatorname{sen} tx$, donde las funciones $\operatorname{sen} tx$, $\cos tx$ son continuas y acotadas; y tenemos $F_n \Rightarrow F(x)$ (porque se trata de funciones de distribución), obtenemos (6.23) aplicando el teorema 6.4 de Helly.

Supongamos ahora que se verifica la condición (6.23). En virtud del teorema 6.3 de Helly, la sucesión $\{F_n(x)\}$ contiene una cierta subsucesión $\{F_{n_k}(x)\}$, que converge débilmente a una cierta función $F(x)$ no decreciente, que verifica $0 \leq F(x) \leq 1$, y que elegimos continua por la derecha. Demostremos que $F(x)$ es una función de distribución. Para ésto hay que demostrar que $F(-\infty) = 0$ y $F(+\infty) = 1$, lo que equivale a demostrar que $\delta = F(+\infty) - F(-\infty) = 1$.

Supongamos que $\delta < 1$ y sea ε tal que $0 < \varepsilon < 1 - \delta$. Como $f(0) = 1$ y $f(t)$ es continua, para γ suficientemente pequeño es válida la desigualdad

$$\frac{1}{2\gamma} \int_{-\gamma}^{\gamma} f(t) dt > 1 - \frac{\varepsilon}{2} > \delta + \frac{\varepsilon}{2}. \quad (6.24)$$

Como $F_{n_k} \rightarrow F$, podemos elegir un real $A > 4/(\gamma\varepsilon)$ que verifique: $F(x)$ es continua en A y en $-A$; para todo k suficientemente grande $\delta_k(A) = F_{n_k}(A) - F_{n_k}(-A) < \delta + \varepsilon/4$.

Introducimos ahora la función

$$B(x) = \int_{-\gamma}^{\gamma} e^{itx} dt = \frac{2}{x} \operatorname{sen}(\gamma x)$$

que, como $|e^{itx}| \leq 1$, verifica $|B(x)| \leq 2\gamma$. Además, como $|\operatorname{sen}(\gamma x)| \leq 1$, tenemos $|B(x)| \leq 2/A$, si $|x| \geq A$.

Cambiando el orden de integración y utilizando la función $B(x)$ recién introducida, tenemos

$$\int_{-\gamma}^{\gamma} f_{n_k}(t) dt = \int_{-\gamma}^{\gamma} \left(\int_{-\infty}^{\infty} e^{itx} dF_{n_k}(x) \right) dt = \int_{-\infty}^{\infty} B(x) dF_{n_k}(x).$$

Partiendo el intervalo de integración, obtenemos la siguiente acotación:

$$\begin{aligned} \left| \int_{-\gamma}^{\gamma} f_{n_k}(t) dt \right| &\leq \left| \int_{\{|x| \leq A\}} B(x) dF_{n_k}(x) \right| \\ &\quad + \left| \int_{\{|x| > A\}} B(x) dF_{n_k}(x) \right| \leq 2\gamma\delta + \frac{2}{A}. \end{aligned}$$

En vista de la elección de A , dividiendo por 2γ , obtenemos

$$\frac{1}{2\gamma} \left| \int_{-\gamma}^{\gamma} f_{n_k}(t) dt \right| \leq \delta + \frac{\varepsilon}{2}.$$

Como $f_{n_k}(t) \rightarrow f(t)$ ($n \rightarrow \infty$) con $|f_{n_k}(t)| \leq 1$, tomando límite en la desigualdad anterior, obtenemos

$$\frac{1}{2\gamma} \left| \int_{-\gamma}^{\gamma} f(t) dt \right| \leq \delta + \frac{\varepsilon}{2},$$

lo que contradice la desigualdad (6.24). Entonces $\delta = 1$ y $F(x)$ es una función de distribución.

Por último observemos, que como $F_{n_k} \rightarrow F$, aplicando la primera parte del teorema, obtenemos que $f(t)$ es la función característica que corresponde a esta distribución $F(x)$.

Para terminar, resta demostrar que toda la sucesión $\{F_n(x)\}$ converge débilmente a $F(x)$. Supongamos que esto no es cierto. Existe entonces una subsucesión $\{F_{n'}(x)\}$, que converge débilmente a una cierta función $G(x)$, que es distinta de $F(x)$, en por lo menos un punto de continuidad. Una argumentación similar a la aplicada a $\{F_{n_k}(x)\}$ nos permite obtener, que $G(x)$ es una función de distribución, y que $f(t)$ es su función característica. Aplicando el teorema 6.2 de unicidad de funciones características obtenemos que $F(x) = G(x)$ para todo x real, contradiciendo nuestro supuesto. Entonces $F_n \rightarrow F$, lo que concluye la demostración. \square

En el capítulo 7 estudiaremos distintas variantes del *teorema central del límite*, cuyas demostraciones se basan el teorema recién demostrado.

6.5. Ejercicios

1. Hallar la función característica de una variable aleatoria: (a) con distribución uniforme en el intervalo $(-\ell, \ell)$; (b) con densidad dada por $p(x) = (1 - \cos x)/(\pi x^2)$.
2. Dada una variable aleatoria X , la variable aleatoria simetrizada, designada X^s , se define mediante la igualdad $X^s = X - Y$, donde Y es una variable aleatoria independiente de X , y con su misma distribución. Demostrar que si X tiene función característica $f(t)$, entonces, X^s tiene función característica $|f(t)|^2$.
3. Consideremos una función característica $f(t)$. Demostrar la desigualdad

$$1 - |f(2t)|^2 \leq 4(1 - |f(t)|^2),$$

válida para todo t real.

4. Consideremos funciones características $f_1(t), \dots, f_n(t)$, y constantes positivas b_1, \dots, b_n , que verifican $b_1 + \dots + b_n = 1$. Demostrar que $b_1 f_1(t) + \dots + b_n f_n(t)$ es una función característica.

5. Determinar si las siguientes son funciones características: (a) $\sin t$; (b) $\cos t$; (c) $\cos^2 t$; (d) $\sin t + \cos t$; (e) $(e^{it} + e^{2it})^3/8$; (f) $\operatorname{Re} f(t)$, donde $f(t)$ es una función característica; (g) $\operatorname{Im} f(t)$ donde $f(t)$ es una función característica.

6. Calcular la varianza $\operatorname{var} X$, de una variable aleatoria X , con función característica $f(t) = (1 + e^{3it})^2/4$.

7. Sea X una variable aleatoria con distribución discreta. Esta distribución se denomina *látice*, si existen dos reales $h > 0$ y a , tales que se verifica $\sum_{k=-\infty}^{\infty} \mathbf{P}(X = a + hk) = 1$. (a) Encontrar una distribución discreta que no sea *látice*. (b) Demostrar que una distribución con función característica $f(t)$ es *látice* si y solo si existe $t_0 \neq 0$ tal que $|f(t_0)| = 1$.

8. Sea X una variable aleatoria con distribución *látice*, que toma los valores $a + hk$ ($k = 0, \pm 1, \pm 2, \dots$), con probabilidades $p_k = \mathbf{P}(X = a + kh)$. Demostrar que se verifica

$$p_k = \frac{h}{2\pi} \int_{\{|t| < \frac{\pi}{h}\}} e^{-it(a+kh)} f(t) dt$$

para todo k entero, donde $f(t)$ es la función característica de X .

9. Utilizando el teorema 6.2 de unicidad, demostrar: si X e Y son variables aleatorias independientes, con distribución de Poisson con parámetros λ_1 y λ_2 respectivamente, entonces $X + Y$ tiene distribución de Poisson con parámetro $\lambda_1 + \lambda_2$.

10. Consideremos una función característica $f(t)$ y dos constantes b, c , que verifican $0 < c < 1$, $b > 0$. Demostrar que si $|f(t)| \leq c$, cuando $|t| \geq b$, entonces $|f(t)| \leq 1 - (1 - c^2)t^2/(8b^2)$, si $|t| < b$.

11. Demostrar que si una función característica verifica la condición

$$\limsup_{|t| \rightarrow \infty} |f(t)| < 1$$

(denominada *condición (C) de Cramér*), entonces, para todo $\varepsilon > 0$, existe un real positivo $c < 1$, tal que $|f(t)| \leq c$, cuando $|t| \geq \varepsilon$. Sugerencia: Utilizar el ejercicio 10.

12. Consideremos una variable aleatoria con función característica $f(t)$. Demostrar que si la variable aleatoria es absolutamente continua, entonces $f(t) \rightarrow 0$, si $|t| \rightarrow \infty$. (Sugerencia: utilizar el teorema de Riemann-Lebesgue).

13. Sea $F(x)$ una función de distribución con función característica $f(t)$. Demostrar la igualdad

$$\int_{-\infty}^{\infty} \frac{x^2}{1+x^2} dF(x) = \int_0^{\infty} e^{-t}(1 - \operatorname{Re} f(t)) dt.$$

14. Una función característica se denomina *infinitamente divisible* si para cada natural $n \geq 1$, existe una función característica $f_n(t)$, tal que $f(t) = (f_n(t))^n$. Demostrar las siguientes proposiciones: (a) Si $f(t)$ y $g(t)$ son funciones características infinitamente divisibles, entonces, $f(t)g(t)$ es una función característica infinitamente divisible. (b) Si $f(t)$ es función característica infinitamente divisible, entonces, $f(t) \neq 0$ para todo t real. (Sugerencia: utilizar la desigualdad del ejercicio 3.)

15. Si una función característica verifica $f(t) = 1 + o(t^2)$ ($t \rightarrow 0$), entonces $f(t) = 1$ para todo t real.

16. Sea $f(t)$ la función característica de una variable aleatoria con distribución no degenerada. Demostrar que existen reales positivos δ y ε , tales que $|f(t)| \leq 1 - \varepsilon t^2$ para $|t| \leq \delta$.

17. Sea X una variable aleatoria con función característica $f(t)$, con distribución látilice, que toma los valores $a + hk$ ($k = 0, \pm 1, \pm 2, \dots$), donde $h > 0$ y a son números reales fijos. El número h se denomina el *paso* de la distribución. El paso h se denomina *maximal*, si no existe un par h_1, a_1 , con $h_1 > h$, tal que $\sum_{k=-\infty}^{\infty} \mathbf{P}(X = a_1 + h_1 k) = 1$. Demostrar que el paso h es maximal si y solo si se verifican las condiciones $|f(2\pi/h)| = 1$, y $|f(t)| < 1$ para $0 < |t| < 2\pi/h$.

18. Una función de distribución, o su función característica $f(t)$, se denomina *estable*, cuando para todo par a_1 y a_2 de números reales positivos,

existen $a > 0$ y b reales, tales que $f(a_1t)f(a_2t) = e^{ibt}f(at)$. (a) Demostrar que esta definición es equivalente a la siguiente: la función de distribución $F(x)$ es estable, si para todo $a_1 > 0$, $a_2 > 0$, b_1 y b_2 , existen $a > 0$ y b reales tales que $F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b)$, donde $*$ designa la convolución. (b) Determinar si son estables las siguientes distribuciones: (i) degenerada, (ii) normal, (iii) uniforme, (iv) binomial, (v) de Poisson.

19. (a) Hallar la función característica de una variable aleatoria con distribución Gama, con parámetros (α, β) . (b) Demostrar que si T_1, \dots, T_n son variables aleatorias independientes con distribución común exponencial de parámetro α , entonces, su suma $T_1 + \dots + T_n$, tiene densidad dada por $p(x) = \alpha^n x^{n-1} e^{-\alpha x} / (n-1)!$.

Capítulo 7

Teorema central del límite

Se denomina *teorema central del límite* a cualquier proposición que establece, bajo determinadas condiciones, que la función de distribución de la suma de una cantidad creciente a infinito de variables aleatorias, converge a la función de distribución normal. Aplicando el teorema central del límite podemos aproximar la distribución de la suma de un gran número de variables aleatorias, mediante la distribución normal. En este capítulo, dedicado a estudiar diversas variantes del teorema central del límite, comenzamos por el teorema de Lindeberg–Lévy que considera sumas de variables independientes e idénticamente distribuidas. En la sección 2 se demuestra un resultado más general: el teorema de Lindeberg, en el cual no se supone que las variables aleatorias consideradas tienen la misma distribución.

7.1. Teorema de Lindeberg–Lévy

Decimos que X_1, X_2, \dots es una *sucesión de variables aleatorias independientes*, cuando las variables aleatorias X_1, \dots, X_n son mutuamente independientes para cada $n = 1, 2, \dots$. Recordemos, que X_1, X_2, \dots es una *sucesión de variables aleatorias idénticamente distribuidas* cuando todas las variables aleatorias consideradas tienen la misma distribución.

Teorema 7.1 (Lindeberg–Lévy).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes e idénticamente distribuidas, con esperanza matemática $\mathbf{E} X_1 = a$

y varianza $\mathbf{var} X_1 = \sigma^2 > 0$. Designemos

$$F_n(x) = \mathbf{P} \left(\frac{X_1 + \cdots + X_n - na}{\sigma\sqrt{n}} \leq x \right).$$

Entonces,

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real,} \quad (7.1)$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ es la distribución normal estándar.

Demostración. La primer etapa consiste en demostrar el teorema en el caso particular en el que $\mathbf{E} X_1 = a = 0$ y $\mathbf{var} X_1 = \sigma^2 = 1$.

Consideremos entonces para cada $n = 1, 2, \dots$ la variable aleatoria

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n).$$

Como $a = 0$ y $\sigma^2 = 1$, tenemos $\mathbf{P}(Z_n \leq x) = F_n(x)$. La demostración se basa en la aplicación del teorema 6.5. Calculemos $f_n(t)$, la función característica de Z_n , en términos de $v(t)$, la función característica de X_1 :

$$\begin{aligned} f_n(t) &= \mathbf{E} e^{itZ_n} = \mathbf{E} e^{i(t/\sqrt{n})\sum_{k=1}^n X_k} = \mathbf{E} \prod_{k=1}^n e^{i(t/\sqrt{n})X_k} \\ &= \prod_{k=1}^n \mathbf{E} e^{i(t/\sqrt{n})X_k} = \left[v\left(\frac{t}{\sqrt{n}}\right) \right]^n, \end{aligned} \quad (7.2)$$

donde utilizamos que las variables aleatorias son idénticamente distribuidas en la última igualdad, y que son independientes en la ante última. Como $\alpha_2 = \mathbf{var} X_1 = 1 < \infty$, aplicando el desarrollo de Taylor (6.10) de orden $k = 2$ para la función característica de X_1 , tenemos

$$v(u) = 1 - \frac{u^2}{2} + o(u^2) \quad (u \rightarrow 0), \quad (7.3)$$

dado que $\alpha_1 = \mathbf{E} X_1 = 0$. Consideremos un real t arbitrario y fijo. Queremos calcular $\lim_{n \rightarrow \infty} f_n(t)$. Si en (7.3) ponemos $u = t/\sqrt{n}$, tenemos

$$v\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \quad (n \rightarrow \infty),$$

dado que $1/n \rightarrow 0$ si y solo si $u \rightarrow 0$.

Verifiquemos ahora la validez de la identidad

$$\ln(1+z) = z + r(z), \quad (7.4)$$

donde $|r(z)| \leq 2|z|^2$, si z es un número complejo que verifica $|z| < 1/2$. En efecto, considerando el desarrollo de Taylor de la función logaritmo, tenemos

$$r(z) = \ln(1+z) - z = \sum_{m=2}^{\infty} (-1)^{m-1} z^m / m.$$

Acotando y sumando la serie geométrica que se obtiene, tenemos

$$|r(z)| \leq \sum_{m=2}^{\infty} |z|^m = |z|^2 \sum_{m=0}^{\infty} |z|^m = \frac{|z|^2}{1-|z|} \leq 2|z|^2,$$

porque $(1-|z|)^{-1} \leq 2$, dado que $|z| \leq 1/2$. Esto prueba (7.4).

Si $z = v(t/\sqrt{n}) - 1 = -t^2/(2n) + o(1/n)$, para n suficientemente grande podemos aplicar la fórmula (7.4), para obtener

$$\ln v\left(\frac{t}{\sqrt{n}}\right) = -\frac{t^2}{2n} + o\left(\frac{1}{n}\right),$$

porque, en este caso $|r(z)| \leq 2|z|^2 = t^4/(2n^2) + o(1/n^2)$.

Estamos en condiciones de tomar logaritmo en la fórmula (7.2):

$$\ln f_n(t) = n \ln v\left(\frac{t}{\sqrt{n}}\right) = n \left[-\frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right] = -\frac{t^2}{2} + o(1).$$

En otras palabras, $f_n(t) \rightarrow e^{-t^2/2}$ si $n \rightarrow \infty$. Como $f(t) = e^{-t^2/2}$ es la función característica de la distribución normal $\Phi(x)$, aplicando el teorema 6.5 obtenemos que $F_n(x) \rightarrow \Phi(x)$ para todo x real (dado que $\Phi(x)$ es continua en \mathbb{R}), concluyendo la primer etapa de la demostración ($a = 0, \sigma^2 = 1$).

Supongamos ahora que $\mathbf{E} X_1 = a, \mathbf{var} X_1 = \sigma^2$, con $a, \sigma > 0$ arbitrarios. Consideremos las variables aleatorias auxiliares $Y_n = (X_n - a)/\sigma$ ($n = 1, 2, \dots$). Es fácil de ver que $\{Y_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, con $\mathbf{E} Y_1 = 0$, y $\mathbf{var} Y_1 = 1$. Entonces

$$\begin{aligned} F_n(x) &= \mathbf{P} \left(\frac{X_1 + \dots + X_n - na}{\sigma\sqrt{n}} \leq x \right) \\ &= \mathbf{P} \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \leq x \right) \rightarrow \Phi(x), \end{aligned}$$

para todo x real si $n \rightarrow \infty$, dado que $\{Y_n\}$ verifica las condiciones de la primer etapa de la demostración. Esto es la tesis del teorema. \square

Observación. Es posible demostrar que la convergencia en (7.1) es uniforme en el conjunto de los x reales. No es difícil verificar esta afirmación directamente; es consecuencia del *teorema de Pólya*: si una sucesión de funciones de distribución $\{G_n(x)\}$ converge a una función de distribución continua $G(x)$ para todo x , entonces, esta convergencia es uniforme en la recta real (ver ejercicio 12, capítulo 5).

Veamos que el teorema límite integral de De Moivre–Laplace de la sección 2.3, es un corolario del teorema de Lindeberg–Lévy recién demostrado.

Consideremos entonces una serie de n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso), y probabilidad de éxito igual a p en cada experimento ($0 < p < 1$). Sea μ la cantidad de éxitos en n experimentos. Veamos como se formula el teorema límite integral de De Moivre–Laplace en términos de variables aleatorias. Para ésto consideremos una sucesión de variables aleatorias X_1, X_2, \dots , cada una de las cuales toma el valor 1 con probabilidad p (si ocurre un éxito) y el valor 0 con probabilidad $q = 1 - p$ (si ocurre un fracaso). Tenemos $\mathbf{E} X_k = p$, $\mathbf{var} X_k = pq > 0$ para cada $i = 1, 2, \dots$. Además, $\mu = \sum_{k=1}^n X_k$, porque la suma contiene tantos sumandos iguales a uno como éxitos ocurren en los primeros n experimentos, siendo nulos los sumandos restantes. La sucesión $\{X_n\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, por lo que es aplicable el teorema de Lindeberg–Lévy. De (7.1) obtenemos, que

$$\mathbf{P} \left(\frac{\mu - np}{\sqrt{npq}} \leq x \right) - \Phi(x) \rightarrow 0 \quad (n \rightarrow \infty), \quad (7.5)$$

uniformemente en el conjunto de los x reales, en vista de la última observación. Poniendo entonces en (7.5) primero $x = b$, luego $x = a$, y restando, obtenemos

$$\mathbf{P} \left(a < \frac{\mu - np}{\sqrt{npq}} \leq b \right) - \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx \rightarrow 0 \quad (n \rightarrow \infty),$$

uniformemente, en el conjunto de los reales $a < b$, que es el contenido del teorema límite integral de De Moivre–Laplace.

7.2. Teorema de Lindeberg

Teorema 7.2 (Lindeberg).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes, con esperanzas a_1, a_2, \dots y varianzas $\sigma_1^2, \sigma_2^2, \dots$, no todas nulas. Designemos

$$V_n(x) = \mathbf{P}(X_n \leq x), \quad B_n = \sum_{k=1}^n \sigma_k^2,$$

$$F_n(x) = \mathbf{P}\left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x\right).$$

Supongamos que se verifica la condición de Lindeberg: Para todo $\varepsilon > 0$

$$\Lambda_n(\varepsilon) = \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x - a_k)^2 dV_k(x) \rightarrow 0 \quad (n \rightarrow \infty). \quad (7.6)$$

Entonces

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real.}$$

La demostración del teorema de Lindeberg utiliza el siguiente resultado de cálculo, que incluimos por conveniencia del lector.

Lema 7.1. Vale la desigualdad

$$\left| e^{ix} - \sum_{\nu=0}^{k-1} \frac{(ix)^\nu}{\nu!} \right| \leq \frac{1}{k!} |x|^k, \quad (7.7)$$

para todo real x , y todo natural $k \geq 1$.

Demostración del lema 7.1. Como

$$\int_0^x e^{it} dt = \frac{1}{i} (e^{ix} - 1)$$

obtenemos que $|e^{ix} - 1| \leq |x|$, demostrando la fórmula (7.7) para $k = 1$.

Para demostrar la validez de (7.7) para $k + 1$, a partir de su validez para k , escribimos

$$I = \int_0^x \left(e^{it} - \sum_{\nu=0}^{k-1} \frac{(it)^\nu}{\nu!} \right) dt = \frac{1}{i} \left(e^{ix} - \sum_{\nu=0}^k \frac{(ix)^\nu}{\nu!} \right).$$

Entonces

$$\left| e^{ix} - \sum_{\nu=0}^k \frac{(ix)^\nu}{\nu!} \right| = |I| \leq \int_0^{|x|} \frac{t^k}{k!} dt = \frac{|x|^{k+1}}{(k+1)!},$$

concluyendo la demostración. \square

Demostración del teorema 7.2 de Lindeberg. Observemos en primer lugar que podemos suponer $a_n = 0$ ($n = 1, 2, \dots$). El caso general en el que las variables aleatorias tienen esperanzas arbitrarias, se reduce a este caso particular mediante la consideración de las variables aleatorias $Y_n = X_n - a_n$ ($n = 1, 2, \dots$) que verifican $\mathbf{E} Y_n = 0$, $\mathbf{var} Y_n = \sigma_n^2$.

Supongamos entonces que $\mathbf{E} X_n = 0$ ($n = 1, 2, \dots$). En primer lugar, demostremos que la condición (7.6) de Lindeberg, que en nuestro caso es: Para todo $\varepsilon > 0$

$$\Lambda_n(\varepsilon) = \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) \rightarrow 0 \quad (n \rightarrow \infty),$$

implica la condición:

$$\frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty). \quad (7.8)$$

En efecto, para $\varepsilon > 0$ arbitrario y para cada $k = 1, 2, \dots, n$, tenemos

$$\begin{aligned} \sigma_k^2 &= \int_{\{|x| < \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) + \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x) \\ &\leq \varepsilon^2 B_n + \sum_{k=1}^n \int_{\{|x| \geq \varepsilon \sqrt{B_n}\}} x^2 dV_k(x). \end{aligned}$$

Entonces, como la cota obtenida no depende de k , dividiendo por B_n tenemos

$$\frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \leq \varepsilon^2 + \Lambda_n(\varepsilon) < \varepsilon^2 + \varepsilon,$$

para n suficientemente grande. Como $\varepsilon > 0$ es arbitrario, hemos demostrado (7.8).

Consideremos ahora $Z_n = \sum_{k=1}^n X_k / \sqrt{B_n}$ y calculemos su función característica $f_n(t)$ en términos de $v_k(t)$, la función característica de X_k ($k =$

$1, \dots, n$). Tenemos

$$\begin{aligned} f_n(t) &= \mathbf{E} e^{itZ_n} = \mathbf{E} e^{i(t/\sqrt{B_n})\sum_{k=1}^n X_k} = \mathbf{E} \prod_{k=1}^n e^{i(t/\sqrt{B_n})X_k} \\ &= \prod_{k=1}^n \mathbf{E} e^{i(t/\sqrt{B_n})X_k} = \prod_{k=1}^n v_k\left(\frac{t}{\sqrt{B_n}}\right), \end{aligned} \quad (7.9)$$

donde utilizamos que las variables aleatorias son independientes.

Para la demostración del teorema es suficiente verificar que $f_n(t) \rightarrow e^{-t^2/2}$ ($n \rightarrow \infty$) y aplicar el teorema 6.5. Para demostrar entonces la convergencia de las funciones características, tomamos logaritmo en (7.9) y utilizamos el siguiente resultado.

Lema 7.2. *Para cada t real, tiene lugar la igualdad*

$$\ln f_n(t) = \sum_{k=1}^n \ln v_k\left(\frac{t}{\sqrt{B_n}}\right) = \sum_{k=1}^n \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right] + R_n(t),$$

donde $R_n(t) \rightarrow 0$ ($n \rightarrow \infty$).

Demostración del lema 7.2. Consideremos

$$\begin{aligned} r_k(t) &= \ln v_k\left(\frac{t}{\sqrt{B_n}}\right) - \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right] \text{ para } k = 1, \dots, n, \\ R_n(t) &= \sum_{k=1}^n r_k(t). \end{aligned}$$

Como $\mathbf{E} X_k = \int_{-\infty}^{\infty} x dV_k(x) = 0$, aplicando el lema 7.1 con $k = 2$, tenemos

$$\begin{aligned} \left| v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right| &= \left| \int_{-\infty}^{+\infty} \left(e^{\frac{itx}{\sqrt{B_n}}} - 1 - \frac{itx}{\sqrt{B_n}} \right) dV_k(x) \right| \\ &\leq \frac{t^2}{2B_n} \int_{-\infty}^{+\infty} x^2 dV_k(x) = \frac{t^2 \sigma_k^2}{2B_n} \end{aligned} \quad (7.10)$$

$$\leq \frac{t^2}{2} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty), \quad (7.11)$$

según vimos en la fórmula (7.8). Luego, si n es suficientemente grande, designando $z_k = v_k(t/\sqrt{B_n}) - 1$, se verifica $|z_k| < 1/2$ para todo $k = 1, \dots, n$, y podemos utilizar el desarrollo del logaritmo (7.4), para obtener

$$|r_k(t)| \leq 2 \left| v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 \right|^2 \leq \frac{t^4}{2B_n^2} \sigma_k^4 \leq \frac{t^4 \sigma_k^2}{2B_n} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2,$$

donde utilizamos (7.10). Por ésto,

$$|R_n(t)| = \sum_{k=1}^n |r_k(t)| \leq \frac{t^4}{2} \sum_{k=1}^n \frac{\sigma_k^2}{B_n} \times \frac{1}{B_n} \max_{1 \leq k \leq n} \sigma_k^2 \rightarrow 0 \quad (n \rightarrow \infty),$$

en vista de (7.8), concluyendo la demostración del lema. \square

Resta la etapa final, que consiste en demostrar que

$$\ln f_n(t) + t^2/2 \rightarrow 0 \quad (n \rightarrow \infty). \quad (7.12)$$

Con este fin, introducimos

$$I_k = v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 + \frac{t^2\sigma_k^2}{2B_n} = \int_{-\infty}^{\infty} \left(e^{itx/\sqrt{B_n}} - 1 - \frac{itx}{\sqrt{B_n}} - \frac{(itx)^2}{2B_n} \right) dV_k(x)$$

para cada $k = 1, 2, \dots$. Acotamos I_k partiendo el dominio de integración en las regiones $\{|x| < \varepsilon\sqrt{B_n}\}$ y $\{|x| \geq \varepsilon\sqrt{B_n}\}$, y utilizando el lema 7.1 con $k = 3$ en la primer región, y con $k = 2$ en la segunda. Las acotaciones que obtenemos de dicho lema, son

$$\begin{aligned} \left| e^{iy} - 1 - iy - \frac{(iy)^2}{2} \right| &\leq \frac{1}{6}|y|^3, \\ \left| e^{iy} - 1 - iy - \frac{(iy)^2}{2} \right| &\leq |e^{iy} - 1 - iy| + \frac{1}{2}|y|^2 \leq |y|^2. \end{aligned}$$

Por eso, si $y = ix/\sqrt{B_n}$, tenemos

$$\begin{aligned} |I_k| &\leq \int_{\{|x| < \varepsilon\sqrt{B_n}\}} \frac{|tx|^3}{6B_n^{3/2}} dV_k(x) + \int_{\{|x| \geq \varepsilon\sqrt{B_n}\}} \frac{|tx|^2}{B_n} dV_k(x) \\ &\leq \frac{\varepsilon\sigma_k^2|t|^3}{6B_n} + \frac{t^2}{B_n} \int_{\{|x| \geq \varepsilon\sqrt{B_n}\}} |x|^2 dV_k(x). \end{aligned} \quad (7.13)$$

Estamos en condiciones de concluir la demostración. Utilizando el lema 7.2, tenemos

$$\begin{aligned} \ln f_n(t) + \frac{t^2}{2} &= \sum_{k=1}^n \left[\ln v_k\left(\frac{t}{\sqrt{B_n}}\right) + \frac{t^2\sigma_k^2}{2B_n} \right] \\ &= \sum_{k=1}^n \left[v_k\left(\frac{t}{\sqrt{B_n}}\right) - 1 - \frac{(it)^2\sigma_k^2}{2B_n} \right] + R_n(t) \\ &= \sum_{k=1}^n I_k + R_n(t). \end{aligned}$$

Aplicando ahora la acotación (7.13), tenemos

$$\left| \ln f_n(t) + \frac{t^2}{2} \right| \leq \sum_{k=1}^n |I_k| + R_n(t) \leq \frac{\varepsilon |t|^3}{6} + t^2 \Lambda_n(\varepsilon) + R_n(t).$$

Como el primer sumando es arbitrariamente pequeño, el segundo converge a cero (aplicando la condición de Lindeberg), y el tercero también tiende a cero (según demostramos en el lema 7.2), obtuvimos (7.12). Con la aplicación del teorema 6.5 concluimos la demostración. \square

El teorema 7.1 de Lindeberg–Lévy, resulta ser un corolario del teorema de Lindeberg, recién demostrado. En efecto, en el caso de variables aleatorias con distribución común $V(x)$, esperanza matemática a y variancia $\sigma^2 > 0$, obtenemos que $B_n = n\sigma^2$, y, para $\varepsilon > 0$ arbitrario, se verifica

$$\Lambda_n(\varepsilon) = \frac{1}{\sigma^2} \int_{\{|x-a| \geq \varepsilon \sigma \sqrt{n}\}} (x-a)^2 dV(x) \rightarrow 0 \quad (n \rightarrow \infty)$$

porque $\sigma^2 = \int_{-\infty}^{\infty} (x-a)^2 V(x) < \infty$. En conclusión, si se verifican las hipótesis del teorema de Lindeberg–Lévy, también se verifican las del teorema de Lindeberg; mientras que las tesis de estos dos teoremas, en el caso particular considerado, coinciden.

Volvamos ahora al caso general, en el que las distribuciones no necesariamente son idénticas. Sea $\{X_n\}$ una sucesión de variables aleatorias que verifica las condiciones del teorema 7.2 de Lindeberg. Consideremos

$$X_{nk} = \frac{X_k - a_k}{\sqrt{B_n}} \quad (k = 1, \dots, n).$$

Poniendo $Z_n = \sum_{k=1}^n (X_k - a_k)/\sqrt{B_n}$, tenemos $Z_n = \sum_{k=1}^n X_{nk}$. Demostremos que la condición de Lindeberg implica la condición: Para todo $\varepsilon > 0$

$$\mathbf{P} \left(\max_{1 \leq k \leq n} |X_{nk}| \geq \varepsilon \right) \rightarrow 0 \quad (n \rightarrow \infty). \quad (7.14)$$

La fórmula (7.14) significa que las variables aleatorias son uniformemente “pequeñas”. Veamos su demostración. Dado $\varepsilon > 0$, tenemos

$$\begin{aligned} \mathbf{P}(|X_{nk}| \geq \varepsilon) &= \mathbf{P}(|X_k - a_k| \geq \varepsilon \sqrt{B_n}) = \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} dV_k(x) \\ &\leq \frac{1}{\varepsilon^2 B_n} \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x - a_k)^2 dV_k(x). \end{aligned}$$

Por ésto,

$$\begin{aligned} \mathbf{P} \left(\max_{1 \leq k \leq n} |X_{nk}| \geq \varepsilon \right) &\leq \mathbf{P} \left(\bigcup_{k=1}^n \{|X_{nk}| \geq \varepsilon\} \right) \leq \sum_{k=1}^n \mathbf{P} (|X_{nk}| \geq \varepsilon) \\ &\leq \frac{1}{\varepsilon^2 B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x - a_k)^2 dV_k(x) = \frac{1}{\varepsilon^2} \Lambda_n(\varepsilon), \end{aligned}$$

obteniendo la condición (7.14).

7.3. Teorema de Lyapunov

Teorema 7.3 (Lyapunov).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes, con esperanzas a_1, a_2, \dots y varianzas $\sigma_1^2, \sigma_2^2, \dots$, no todas nulas. Designemos

$$B_n = \sum_{k=1}^n \sigma_k^2, \quad F_n(x) = \mathbf{P} \left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x \right).$$

Supongamos que se verifica la condición de Lyapunov: Existe $\delta > 0$ tal que

$$L_n(\delta) = \frac{1}{B_n^{1+\delta/2}} \sum_{k=1}^n \mathbf{E} |X_n - a_n|^{2+\delta} \rightarrow 0 \quad (n \rightarrow \infty). \quad (7.15)$$

Entonces

$$F_n(x) \rightarrow \Phi(x) \quad (n \rightarrow \infty) \text{ para todo } x \text{ real.}$$

Observemos que la condición de Lyapunov implica que existen los momentos de orden $2 + \delta$ de las variables aleatorias X_k ($k = 1, 2, \dots$).

Demostración. Como la tesis del teorema 7.2 y la del teorema 7.3 coinciden, es suficiente demostrar que la condición (7.15) de Lyapunov implica la condición (7.6) de Lindeberg. En efecto, supongamos que se verifica la

condición (7.15) para un cierto $\delta > 0$. Dado $\varepsilon > 0$ arbitrario, tenemos

$$\begin{aligned}\Lambda_n(\varepsilon) &= \frac{1}{B_n} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} (x-a_k)^2 dV_k(x) \\ &\leq \frac{1}{\varepsilon^\delta B_n^{1+\delta/2}} \sum_{k=1}^n \int_{\{|x-a_k| \geq \varepsilon \sqrt{B_n}\}} |x-a_k|^{2+\delta} dV_k(x) \\ &\leq \frac{1}{\varepsilon^\delta B_n^{1+\delta/2}} \sum_{k=1}^n \int_{-\infty}^{\infty} |x-a_k|^{2+\delta} dV_k(x) = \frac{1}{\varepsilon^\delta} L_n(\delta) \rightarrow 0 \quad (n \rightarrow \infty).\end{aligned}$$

De aquí se obtiene la demostración, resultando el teorema de Lyapunov un caso particular del teorema de Lindeberg. \square

Como conclusión de este capítulo haremos algunas consideraciones relativas a la *velocidad de convergencia* en el teorema central del límite.

Lyapunov demostró que si se verifican las condiciones del teorema 7.3 con $0 < \delta < 1$, entonces, existe una constante C tal que, para n suficientemente grande, se verifica

$$|F_n(x) - \Phi(x)| \leq CL_n(\delta),$$

uniformemente en el conjunto de los x reales.

En el caso $\delta = 1$ Esseen obtuvo la siguiente desigualdad, válida para *todo* natural $n = 1, 2, \dots$. Sean X_1, \dots, X_n variables aleatorias independientes, con esperanzas a_1, \dots, a_n y varianzas $\sigma_1^2, \dots, \sigma_n^2$, no todas nulas. Supongamos que $\mathbf{E}|X_k - a_k|^3 < \infty$ ($k = 1, \dots, n$). Designemos

$$\begin{aligned}B_n &= \sum_{k=1}^n \sigma_k^2, & F_n(x) &= \mathbf{P}\left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n (X_k - a_k) \leq x\right), \\ L_n &= L_n(1) = \frac{1}{B_n^{3/2}} \sum_{k=1}^n \mathbf{E}|X_n - a_n|^3.\end{aligned}$$

Entonces

$$|F_n(x) - \Phi(x)| \leq AL_n \quad \text{para todo } x \text{ real y todo } n = 1, 2, \dots, \quad (7.16)$$

donde $A > 0$ es una constante absoluta. (Está demostrado que la acotación (7.16) es válida con $A = 0,8$). De (7.16) se obtiene que si $L_n \rightarrow 0$ entonces $F_n(x) \rightarrow \Phi(x)$ para todo x real.

La desigualdad (7.16) es válida también si $\mathbf{E}|X_n - a_n|^{2+\delta} < \infty$ ($k = 1, \dots, n$) para algún $0 < \delta \leq 1$, definiendo $L_n(\delta)$ como en (7.15).

En el caso particular en el que X_1, \dots, X_n son variables aleatorias idénticamente distribuidas, con $a = \mathbf{E} X_1$, $\sigma^2 = \mathbf{var} X_1$, y $\delta = 1$, los resultados anteriores son los siguientes. Si designamos

$$L_n = \frac{\rho}{\sqrt{n}}, \text{ con } \rho = \frac{\mathbf{E}|X_1 - a|^3}{\sigma^3},$$

aplicando la desigualdad (7.16) de Esseen, obtenemos

$$|F_n(x) - \Phi(x)| \leq \frac{A\rho}{\sqrt{n}} \quad (7.17)$$

para todo x real y todo $n = 1, 2, \dots$. Es posible demostrar (ver por ejemplo §5.2 en [5]), que sin la introducción de condiciones adicionales, esta acotación es óptima en el siguiente sentido: el término \sqrt{n} en el denominador, a la derecha en (7.17), no se puede sustituir por una función $g(n)$, que verifique $g(n)/\sqrt{n} \rightarrow \infty$ ($n \rightarrow \infty$).

7.4. Ejercicios

1. Al disparar a un blanco, se obtienen 10 puntos con probabilidad 0,3; 9 puntos con probabilidad 0,3; 8 con probabilidad 0,2; 7 con probabilidad 0,1 y 6 con probabilidad 0,1. Utilizando el teorema central del límite, estimar la probabilidad de que, al realizar 100 disparos, se obtengan más de 870 puntos.
2. Se tira un dado 500 veces. Hallar un valor aproximado para la probabilidad de que la suma de puntos obtenidos sea mayor que 1800.
3. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, e idénticamente distribuidas, con $\mathbf{E} X_1 = a$, y $\mathbf{var} X_1 = \sigma^2 > 0$. Verificar que $\{Y_n\}$, con $Y_n = (X_n - a)/\sigma$, es una sucesión de variables aleatorias independientes, idénticamente distribuidas, con esperanza nula, y varianza igual a uno.
4. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifican $\mathbf{P}(X_n = n^a) = \mathbf{P}(X_n = -n^a) = 1/2$ para cada $n = 1, 2, \dots$, donde $a > -1/2$. ¿Es aplicable el teorema central del límite a esta sucesión?

5. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza nula. Supongamos que $|X_n| \leq C$ para todo n , donde C es una cierta constante. Sea $B_n = \sum_{k=1}^n \text{var } X_k$. Demostrar que si $B_n \rightarrow \infty$ ($n \rightarrow \infty$), entonces, es aplicable el teorema central del límite a esta sucesión, es decir $\mathbf{P}\left(\frac{1}{\sqrt{B_n}} \sum_{k=1}^n X_k \leq x\right) \rightarrow \Phi(x)$.

6. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifican $\mathbf{P}(X_n = -1/\sqrt{n}) = \mathbf{P}(X_n = 1/\sqrt{n}) = p$, $\mathbf{P}(X_n = 0) = 1 - 2p$ para todo n , y algún p en el intervalo $0 < p \leq 1/2$. ¿Es aplicable el teorema central del límite a esta sucesión?

7. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tales que, cada variable aleatoria X_n , tiene distribución uniforme, en el intervalo $(-\sqrt{n}, \sqrt{n})$. Demostrar que para esta sucesión, es aplicable el teorema central del límite.

8. Sea μ la cantidad de éxitos, en una serie n experimentos independientes, con dos resultados posibles cada uno (éxito y fracaso). Sea p_k la probabilidad de que ocurra un éxito en el k -ésimo experimento, y $q_k = 1 - p_k$, la probabilidad de que ocurra un fracaso ($k = 1, 2, \dots$). Demostrar que si $\sum_{k=1}^{\infty} p_k q_k = \infty$, entonces, la función de distribución de la variable aleatoria $(\mu - \sum_{k=1}^n p_k) (\sum_{k=1}^n p_k q_k)^{-1/2}$ (cantidad de éxitos, normalizados), converge a la distribución normal estándar, si $n \rightarrow \infty$.

9. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, que verifica la condición de Lindeberg. Demostrar que $B_n \rightarrow \infty$. (Sugerencia: Utilizar, que la condición de Lindeberg implica la condición (7.8).)

10. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza matemática nula. Supongamos que se verifica

$$\sum_{k=1}^n \mathbf{E} |X_k|^3 \leq Bn, \quad \sum_{k=1}^n \mathbf{E} |X_k|^2 \geq An$$

para todo n , donde A y B son constantes positivas. ¿Es aplicable el teorema central del límite a esta sucesión?

11. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, con esperanza matemática nula. Supongamos que se verifica

$$\sum_{k=1}^n \mathbf{E} |X_k|^2 \left| \ln |X_k| \right|^{1+\delta} \leq Bn, \quad \sum_{k=1}^n \mathbf{E} |X_k|^2 \geq An,$$

para todo n y algún $\delta > 0$, donde A y B son constantes positivas. Demostrar que para esta sucesión es aplicable el teorema central del límite. (Sugerencia: verificar la condición de Lindeberg).

12. Consideremos una sucesión $\{X_n\}$ de variables aleatorias independientes, idénticamente distribuidas, con distribución de Poisson con parámetro

1. Hallar $\lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{1}{\sqrt{n}} \left(\sum_{k=1}^n X_k - n \right) \leq x \right)$.

13. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tal que que para cada $n = 1, 2, \dots$, la variable aleatoria X_n tiene distribución normal, con $\mathbf{E} X_n = 0$, y $\mathbf{var} X_n = 2^{2n}$. (a) ¿Es aplicable el teorema central del límite a esta sucesión? (b) ¿Se cumple la condición de Lindeberg para esta sucesión de variables aleatorias?

14. Sea $\Phi(x)$ la distribución normal estándar. Demostrar la desigualdad

$$1 - \Phi(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2},$$

para todo $x > 0$.

15. Demostrar la fórmula

$$1 - \Phi(x) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \left(1 + O\left(\frac{1}{x^2}\right) \right) \quad (x \rightarrow \infty). \quad (7.18)$$

(Sugerencia: Utilizar la identidad $\int_x^\infty e^{-t^2/2} dt = -\int_x^\infty \frac{1}{t} d(e^{-t^2/2})$ e integrar por partes.)

16. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, idénticamente distribuidas, con esperanza nula y varianza $\sigma^2 > 0$. Designemos

$F_n(x) = \mathbf{P} \left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n X_k \leq x \right)$. Demostrar que si se verifica

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}}$$

para todo x real, y todo n natural, donde C es una constante positiva, entonces, para $0 < \varepsilon < 1$, se verifica

$$\frac{1 - F_n(x)}{1 - \Phi(x)} \rightarrow 1 \quad (n \rightarrow \infty),$$

uniformemente, en el conjunto $0 \leq x \leq (1 - \varepsilon)\sqrt{\ln n}$. (Sugerencia: utilizar la fórmula (7.18).)

Capítulo 8

Cadenas de Markov

Hasta el momento hemos considerado sucesiones de variables aleatorias independientes, casi exclusivamente. Comenzaremos el estudio de las variables aleatorias dependientes considerando las *cadenas de Markov*.

8.1. Definiciones

Consideremos un conjunto \mathbf{I} , finito o numerable. Sea X_0, X_1, X_2, \dots una sucesión de variables aleatorias que toman valores en \mathbf{I} , definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. El conjunto \mathbf{I} se denomina *espacio de estados*, y sus elementos, los *estados*, se designan mediante las letras i, j, k, ℓ , con índices o sin ellos. Si X_0, X_1, X_2, \dots es una sucesión de variables aleatorias independientes, los sucesos

$$\mathbf{A} = \{X_0 = i_0, \dots, X_n = i_n\}, \quad \mathbf{B} = \{X_{n+1} = i_{n+1}, \dots, X_{n+m} = i_{n+m}\}$$

son independientes para cualquier $n = 0, 1, \dots$, cualquier $m = 1, 2, \dots$, y cualquier sucesión de estados i_0, \dots, i_{n+m} , como surge de aplicar la proposición 3.2. La dependencia que estudiaremos, llamada *dependencia markoviana*, consiste en que la probabilidad del suceso \mathbf{B} depende únicamente del valor que toma la variable aleatoria X_n , y no de los valores que toman las variables aleatorias X_0, \dots, X_{n-1} . Si el índice de la sucesión representa el tiempo y n es el instante presente en una cadena de Markov, podemos decir: la probabilidad de un suceso en el futuro, que ocurre en los instantes $n+1, \dots, n+m$, depende solamente del estado en que se encuentra la sucesión en el instante presente n , y no de los estados en que se encontró en los instantes pasados $0, 1, \dots, n-1$. La definición formal es la siguiente.

Definición 8.1. Consideremos una sucesión X_0, X_1, X_2, \dots de variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, que toman valores en un conjunto \mathbf{I} , finito o numerable (que llamamos espacio de estados).

(a) Decimos que la sucesión dada es una cadena de Markov, si se verifica

$$\mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) = \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n), \quad (8.1)$$

para todo $n = 1, 2, \dots$, y cualquier sucesión de estados i_0, \dots, i_{n+1} en \mathbf{I} , siempre que $\mathbf{P}(X_n = i_n, \dots, X_0 = i_0) > 0$. La identidad (8.1) se llama propiedad de Markov.

(b) Decimos que una cadena de Markov es homogénea en el tiempo cuando para todo par de estados i, j la probabilidad condicional $\mathbf{P}(X_{n+1} = j | X_n = i)$ no depende de n . Es decir, cuando se verifica

$$\begin{aligned} \mathbf{P}(X_1 = j | X_0 = i) &= \mathbf{P}(X_2 = j | X_1 = i) \\ &= \dots = \mathbf{P}(X_{n+1} = j | X_n = i). \end{aligned}$$

En general, decimos cadena de Markov, para referirnos a una cadena de Markov homogénea en el tiempo.

Consideremos una cadena de Markov X_0, X_1, X_2, \dots con espacio de estados \mathbf{I} , y dos estados i, j . Designamos

$$p_{ij} = \mathbf{P}(X_1 = j | X_0 = i), \quad \pi_i = \mathbf{P}(X_0 = i).$$

La matriz $\mathbf{P} = (p_{ij})_{i \in \mathbf{I}, j \in \mathbf{I}}$ (posiblemente infinita) se denomina *matriz de transición*, y el vector $\pi = (\pi_i)_{i \in \mathbf{I}}$ *distribución inicial* de la cadena de Markov. Es sencillo de ver que la matriz de transición verifica las siguientes propiedades¹:

(M1) $p_{ij} \geq 0$ para todo par de estados i, j en \mathbf{I} ;

(M2) $\sum_j p_{ij} = 1$ para todo estado i en \mathbf{I} .

Por su parte, la distribución inicial verifica las propiedades:

(D1) $\pi_i \geq 0$ para todo estado i en \mathbf{I} ;

¹Cuando no sea estrictamente necesario omitiremos el espacio de estados \mathbf{I} en la notación, escribiendo, por ejemplo, (π_i) en lugar de $(\pi_i)_{i \in \mathbf{I}}$, ó $\sum_i p_{ij}$ en lugar de $\sum_{i \in \mathbf{I}} p_{ij}$.

$$(D2) \sum_i \pi_i = 1.$$

Ejemplo 8.1. Cadenas de Markov finitas. Prestamos especial atención a las cadenas de Markov cuyo espacio de estados es un conjunto finito, que denominamos *cadenas de Markov finitas*. Si por ejemplo tenemos $\mathbf{I} = \{0, 1, \dots, d\}$, entonces, la matriz de transición \mathbf{P} es

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0d} \\ p_{10} & p_{11} & \cdots & p_{1d} \\ \vdots & \vdots & & \vdots \\ p_{d0} & p_{d1} & \cdots & p_{dd} \end{bmatrix}$$

y la distribución inicial $\pi = (\pi_0, \dots, \pi_d)$. En el caso particular en el que $p_{ij} = \pi_j$ ($i, j = 1, \dots, d$) obtenemos una sucesión de variables aleatorias independientes como en el ejemplo 8.2, cuando el espacio de estados que allí se considera es finito.

Ejemplo 8.2. Variables aleatorias independientes.

Consideremos una sucesión X_0, X_1, X_2, \dots de variables aleatorias independientes, idénticamente distribuidas, que toman valores en el conjunto de los números enteros $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, con distribución de probabilidad dada por $\mathbf{P}(X_0 = i) = \pi_i$ ($i \in \mathbb{Z}$). Observemos que esta sucesión es una cadena de Markov, ya que tenemos

$$\begin{aligned} \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) &= \mathbf{P}(X_{n+1} = i_{n+1}) \\ &= \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n). \end{aligned}$$

La distribución inicial es $\pi = (\pi_i)$, y como $\mathbf{P}(X_1 = j | X_0 = i) = \pi_j$, la matriz de transición $\mathbf{P} = (p_{ij})$ está dada por $p_{ij} = \pi_j$ ($i \in \mathbb{Z}, j \in \mathbb{Z}$).

Ejemplo 8.3. Paseos al azar.

Consideremos ahora una sucesión X_1, X_2, \dots de variables aleatorias independientes, idénticamente distribuidas, que toman valores enteros, con probabilidades $\mathbf{P}(X_1 = i) = \delta_i$ ($i \in \mathbb{Z}$). Las sumas parciales de la sucesión considerada se definen mediante

$$S_0 = 0, \quad S_n = X_1 + \cdots + X_n \quad (n = 1, 2, \dots).$$

La sucesión de sumas S_0, S_1, S_2, \dots se llama *paseo al azar*. Como

$$\begin{aligned} \mathbf{P}(S_{n+1} = i_{n+1} | S_n = i_n, \dots, S_0 = i_0) \\ &= \mathbf{P}(S_n + X_{n+1} = i_{n+1} | S_n = i_n, \dots, S_0 = i_0) \\ &= \mathbf{P}(S_n + X_{n+1} = i_{n+1} | S_n = i_n) = \pi_{i_{n+1} - i_n}, \end{aligned}$$

un paseo al azar es una cadena de Markov, homogénea en el tiempo. Luego, la sucesión S_0, S_1, S_2, \dots es una cadena de Markov con distribución inicial $\pi = (\pi_i)$ concentrada en el origen (es decir, $\pi_0 = 1$, $\pi_i = 0$ si $i \neq 0$), y matriz de transición $\mathbf{P} = (p_{ij})$ dada por $p_{ij} = \delta_{j-i}$ ($i \in \mathbb{Z}, j \in \mathbb{Z}$). Cuando las probabilidades de transición de i a j dependen, únicamente, de la diferencia $j - i$ como en este caso, decimos que la cadena de Markov es *homogénea en el espacio*.

En el caso particular en el que $\delta_1 = p$ ($0 < p < 1$), $\delta_{-1} = 1 - p$, y $\delta_i = 0$, si $i \neq \pm 1$, tenemos un *paseo al azar simple*.

Los ejemplos anteriores muestran que la definición de cadena de Markov incluye, como caso particular, a las sucesiones de variables aleatorias independientes e idénticamente distribuidas, y también a las sucesiones formadas por sumas parciales de variables aleatorias independientes e idénticamente distribuidas, cuando las variables aleatorias toman valores enteros.

Consideremos ahora las *probabilidades de transición de orden n* de una cadena de Markov X_0, X_1, X_2, \dots con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π . Dados dos estados i, j , designando

$$p_{ij}^n = \mathbf{P}(X_n = j | X_0 = i), \quad \pi_i^n = \mathbf{P}(X_n = i),$$

se conforma una matriz $\mathbf{P}^n = (p_{ij}^n)$, que llamamos *matriz de transición de orden n* , y un vector $\pi^n = (\pi_i^n)$, que llamamos *distribución de probabilidad en el instante n* de la cadena de Markov. Observemos que π^0 es la distribución inicial de la cadena de Markov considerada, \mathbf{P}^1 su matriz de transición, y $\mathbf{P}^0 = (p_{ij}^0)$ es la matriz identidad, es decir

$$p_{ii}^0 = 1, \quad p_{ij}^0 = 0, \quad \text{si } i \neq j.$$

Las probabilidades de transición verifican la *ecuación de Kolmogorov-Chapman*:

$$p_{ij}^{m+n} = \sum_k p_{ik}^m p_{kj}^n, \quad (8.2)$$

para todo par de índices m, n y todo par de estados i, j ; en notación matricial la ecuación (8.2) es

$$\mathbf{P}^{m+n} = \mathbf{P}^m \times \mathbf{P}^n, \quad (8.3)$$

donde \times designa el producto de matrices. En efecto, aplicando la fórmula de la probabilidad total y la propiedad de Markov (8.1), se obtiene

$$\begin{aligned} p_{ij}^{m+n} &= \mathbf{P}(X_{m+n} = j | X_0 = i) = \sum_k \mathbf{P}(X_{m+n} = j, X_m = k | X_0 = i) \\ &= \sum_k \mathbf{P}(X_{m+n} = j | X_m = k) \mathbf{P}(X_m = k | X_0 = i) = \sum_k p_{ik}^m p_{kj}^n. \end{aligned}$$

Análogamente se prueba para la distribución de probabilidad en el instante n , que tiene lugar la identidad

$$\pi_j^{m+n} = \sum_k \pi_k^m p_{kj}^n, \quad (8.4)$$

para todo par de índices m, n y todo estado i ; en notación matricial, escribimos

$$\pi^{m+n} = \pi^m \times \mathbf{P}^n.$$

En particular, de (8.3) resulta que $\mathbf{P}^n = \mathbf{P} \times \mathbf{P}^{n-1}$. Esta fórmula aplicada n veces, da como resultado

$$\mathbf{P}^n = \underbrace{\mathbf{P} \times \cdots \times \mathbf{P}}_n.$$

En conclusión, la matriz de transición de orden n es la potencia n -ésima de la matriz de transición \mathbf{P} , y es correcto interpretar el superíndice n en la notación \mathbf{P}^n como la potencia n -ésima de la matriz \mathbf{P} . La distribución de probabilidad en el instante n se obtiene mediante la fórmula

$$\pi^n = \pi \times \mathbf{P}^n, \quad (8.5)$$

que también se escribe

$$\pi_j^n = \sum_k \pi_k p_{kj}^n. \quad (8.6)$$

Calculemos ahora, para una elección de índices n_1, \dots, n_k arbitraria, la distribución del vector aleatorio $(X_{n_1}, \dots, X_{n_k})$, que llamamos *distribución finito-dimensional* de la cadena de Markov. Es claro que para este cálculo es suficiente conocer las probabilidades de la forma

$$\mathbf{P}(X_0 \in A_0, \dots, X_n \in A_n) \quad (n = 0, 1, \dots),$$

donde A_0, \dots, A_n son subconjuntos arbitrarios de \mathbf{I} . A su vez, para calcular estas últimas probabilidades, es suficiente, dada una sucesión de estados i_0, \dots, i_n , conocer las probabilidades

$$\begin{aligned} & \mathbf{P}(X_n = i_n, \dots, X_0 = i_0) \\ &= \mathbf{P}(X_n = i_n | X_{n-1} = i_{n-1}) \cdots \mathbf{P}(X_1 = i_1 | X_0 = i_0) \mathbf{P}(X_0 = i_0) \\ &= p_{i_{n-1}i_n} \cdots p_{i_0i_1} \pi_{i_0}. \end{aligned} \quad (8.7)$$

Este último cálculo muestra que las distribuciones finito-dimensionales de una cadena de Markov se determinan una vez conocida su matriz de transición \mathbf{P} y su distribución inicial π . Mas aún, es posible demostrar que la probabilidad de un suceso, que depende de una cantidad arbitraria y no necesariamente finita de variables aleatorias de la cadena de Markov, está determinada una vez conocidas \mathbf{P} y π , y que, recíprocamente, dado un conjunto \mathbf{I} finito o numerable, una matriz $\mathbf{P} = (p_{ij})_{i \in \mathbf{I}, j \in \mathbf{I}}$ que verifica las propiedades (M1) y (M2) (ver página 172), y un vector $\pi = (\pi_i)_{i \in \mathbf{I}}$, que verifica las propiedades (D1) y (D2), existe un espacio de probabilidad y una cadena de Markov homogénea con espacio de estados \mathbf{I} , con variables aleatorias definidas en este espacio de probabilidad, que tiene a \mathbf{P} como matriz de transición y a π como distribución inicial². De aquí obtenemos una conclusión importante: Las propiedades probabilísticas de una cadena de Markov, es decir, las propiedades que se expresan a través de probabilidades de sucesos, dependen únicamente de la matriz de transición y de la distribución inicial de la cadena de Markov. Por ésto, para estudiar una cadena de Markov es suficiente determinar su espacio de estados, su matriz de transición, y su distribución inicial.

Cuando es necesario explicitar la distribución inicial π de una cadena de Markov, escribimos \mathbf{P}_π en vez de \mathbf{P} . En particular, si esta distribución inicial está concentrada en un estado i , escribimos \mathbf{P}_i en vez de \mathbf{P}_π , y decimos que la cadena de Markov *parte de i* , ya que $\mathbf{P}_i(X_0 = i) = 1$.

Consideremos un suceso arbitrario, de la forma

$$\mathbf{A} = \{X_{n_1} = i_{n_1}, \dots, X_{n_k} = i_{n_k}\}.$$

Observemos que las probabilidades \mathbf{P}_π y \mathbf{P}_i correspondientes a una cadena de Markov con la misma matriz de transición, verifican

$$\mathbf{P}_\pi(\mathbf{A} | X_0 = i) = \mathbf{P}_i(\mathbf{A})$$

²Ver por ejemplo §VIII.1 en Shiryaev [7].

como resulta de poner $\pi_{i_0} = 1$ en la fórmula (8.7), verificándose también

$$\mathbf{P}_\pi(\mathbf{A}) = \sum_i \mathbf{P}_\pi(\mathbf{A} | X_0 = i)\pi_i = \sum_i \mathbf{P}_i(\mathbf{A})\pi_i,$$

como resulta de aplicar la fórmula de la probabilidad total.

Designamos mediante \mathbf{E}_π y \mathbf{E}_i la esperanzas matemáticas con respecto de las probabilidades \mathbf{P}_π y \mathbf{P}_i respectivamente.

8.2. Clasificación de estados. Estados esenciales y periódicos

Consideremos una cadena de Markov con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π . Estamos interesados en estudiar el comportamiento asintótico de esta cadena de Markov. Más precisamente, dados dos estados i, j , consideramos los límites

$$\lim_{n \rightarrow \infty} p_{ij}^n, \quad \lim_{n \rightarrow \infty} \pi_i^n,$$

y nos plantemos las preguntas siguientes: (a) ¿existen dichos límites?; (b) en caso de existir ¿son positivos o nulos?; (c) en caso de ser positivos ¿cómo se calculan? Las respuestas a estas preguntas dependerán del tipo de estados i, j considerados, según las clasificaciones que iremos introduciendo. Comenzamos considerando, en esta sección, clasificaciones de los estados de una cadena de Markov que dependen de las matrices de transición de orden n : estados *esenciales*, *comunicación* entre estados, y período de un estado. Dedicamos la siguiente sección al estudio de la *recurrencia*. En la última sección concluimos el estudio, obteniendo condiciones que permiten dar respuestas a las preguntas planteadas.

Consideremos nuevamente dos estados i, j . Decimos que *de i se llega a j* (o que *de i se accede a j*) y escribimos $i \rightarrow j$, si existe un natural $n \geq 0$ tal que se verifica $p_{ij}^n > 0$. Observemos que esta relación entre estados cumple la *propiedad transitiva*:

$$\text{Si } i \rightarrow j \text{ y } j \rightarrow k, \text{ entonces } i \rightarrow k. \quad (8.8)$$

En efecto, como $i \rightarrow j$, existe r natural tal que $p_{ij}^r > 0$; análogamente, existe s natural tal que $p_{jk}^s > 0$. Aplicando la ecuación (8.2) de Kolmogorov-Chapman, tenemos

$$p_{ik}^{r+s} = \sum_{\ell} p_{i\ell}^r p_{\ell k}^s \geq p_{ij}^r p_{jk}^s > 0,$$

obteniendo que $i \rightarrow k$.

Estamos ahora en condiciones de introducir la primer clasificación de estados. Decimos que un estado i es *esencial* si para todo j , tal que $i \rightarrow j$, se verifica $j \rightarrow i$. En caso contrario, decimos que i es un estado *no esencial*.

Consideremos un estado esencial i de una cadena de Markov. Si para algún estado j tenemos $i \rightarrow j$, entonces j también es esencial. En efecto, consideremos un estado k , tal que $j \rightarrow k$. Aplicando la propiedad transitiva (8.8) tenemos $i \rightarrow k$. Por ser i esencial se cumple $k \rightarrow i$, y podemos aplicar nuevamente la propiedad transitiva para obtener $k \rightarrow j$, concluyendo de esta forma que j es esencial.

Como consecuencia de la proposición anterior resulta que cada vez que una cadena de Markov se encuentra en el conjunto de los estados esenciales, no lo abandona. Por ésto, para el estudio del comportamiento asintótico de una cadena de Markov, consideramos en primera instancia el conjunto de los estados esenciales. El comportamiento asintótico para los estados no esenciales es sencillo, como se ve a continuación de la proposición 8.4.

Consideremos el conjunto de los estados esenciales. Si para un par de estados esenciales i, j tenemos $i \rightarrow j$, entonces $j \rightarrow i$. Decimos entonces que i y j se comunican, y escribimos $i \leftrightarrow j$. Como $p_{ii}^0 = 1$ resulta que todo estado verifica $i \leftrightarrow i$. Además, la definición dada es simétrica: $i \leftrightarrow j$ si y solo si $j \leftrightarrow i$. Por último, en vista de la propiedad (8.8), obtenemos que la comunicación entre estados también verifica la propiedad transitiva: si $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces $i \leftrightarrow k$. En conclusión, la relación \leftrightarrow es una relación de equivalencia. Como consecuencia, el conjunto de los estados esenciales de una cadena de Markov se descompone en una unión disjunta de subconjuntos, que llamamos *clases irreducibles* o simplemente *clases*, con la siguiente propiedad: i y j están en la misma clase si y solo si $i \leftrightarrow j$. Si una clase irreducible está formada por un único estado decimos que el estado es *absorbente*.

Nos interesa destacar el caso en el que existe una única clase: si el espacio de estados es la única clase irreducible decimos que la cadena de Markov es *irreducible*.

Ejemplo 8.4. Paseo al azar simple.

Consideremos un *paseo al azar simple*, es decir, una cadena de Markov con espacio de estados $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ (el conjunto de los números enteros), distribución inicial $\pi = (\pi_i)$ concentrada en el origen, es decir,

$$\pi_0 = 1, \quad \pi_i = 0 \quad \text{si } i \neq 0,$$

y matriz de transición $P = (p_{ij})$, dada por

$$p_{i,i+1} = p, \quad p_{i,i-1} = q, \quad p_{ij} = 0, \quad \text{si } i \neq j \pm 1,$$

donde $p + q = 1$, y $0 < p < 1$ (ver ejemplo 8.3).

Calculemos las probabilidades de transición de orden n . En cada instante pueden ocurrir únicamente dos sucesos: o bien la posición aumenta en una unidad con probabilidad p (y decimos que ocurre un éxito), o bien disminuye en una unidad con probabilidad q (y decimos que ocurre un fracaso). Estamos en presencia entonces de un esquema de Bernoulli, como los considerados en el capítulo 2. Para llegar de i a j en n pasos, tienen que ocurrir r éxitos y s fracasos, con $r + s = n$, y $r - s = j - i$. De aquí, $2r = n + j - i$, y si $n + j - i$ no es un número par, $p_{ij}^n = 0$. Si $n + j - i$ es par, tenemos

$$p_{ij}^n = C_r^n p^r q^{n-r}, \quad r = (n + j - i)/2, \quad (8.9)$$

como resulta de aplicar la proposición 2.1. Obtenemos entonces las siguientes conclusiones: si $j - i$ es par las probabilidades de transición de orden impar son nulas, si $j - i$ es impar son nulas las de orden par, pero siempre tenemos $i \rightarrow j$ y estamos en presencia de una cadena de Markov irreducible.

En el ejemplo anterior se puede observar, que la cantidad de pasos necesaria para retornar a un mismo estado es necesariamente un número par. Esto motiva la siguiente definición.

Definición 8.2 (Período de un estado). *Decimos que un estado i tiene período d , que designamos $d(i)$, si el retorno a i es posible únicamente en un número de pasos múltiplo de d , siendo éste el mayor número natural que verifica esa propiedad. En otras palabras*

$$d(i) = \text{m. c. d.} \{n \geq 1 : p_{ii}^n > 0\}$$

donde m. c. d. es la abreviatura de máximo común divisor. Si $d(i) > 1$, decimos que i es un estado periódico; si $d(i) = 1$, que es aperiódico.

La periodicidad es una *propiedad de clase* en el sentido siguiente.

Proposición 8.1. *Si dos estados se comunican tienen el mismo período. Es decir, $i \leftrightarrow j$ implica $d(i) = d(j)$.*

Demostración. Demostremos primero que cuando $i \leftrightarrow j$, se verifica:

$$\text{Si } p_{jj}^n > 0 \text{ entonces } n \text{ es múltiplo de } d(i). \quad (8.10)$$

En efecto, como $i \rightarrow j$, existe r tal que $p_{ij}^r > 0$; análogamente, existe s tal que $p_{ji}^s > 0$. Luego, aplicando la ecuación (8.2), obtenemos que $p_{ii}^{r+s} \geq p_{ij}^r p_{ji}^s > 0$. En conclusión $r + s$ es múltiplo de $d(i)$. Aplicando ahora dos veces la ecuación (8.2), obtenemos

$$p_{ii}^{r+s+n} \geq p_{ij}^r p_{jj}^n p_{ji}^s > 0,$$

y $r + s + n$ también es múltiplo de $d(i)$; luego n es múltiplo de $d(i)$, probando (8.10).

De la proposición (8.10) resulta que $d(i) \leq d(j)$. Para concluir la demostración observemos que $d(j) \leq d(i)$ resulta de intercambiar i con j . \square

En vista de la proposición anterior obtenemos que todos los estados de una clase irreducible tienen el mismo período. Decimos entonces, por ejemplo, que una clase tiene período d , o que una clase es aperiódica. En particular decimos *cadena de Markov aperiódica*, para referirnos a una cadena de Markov irreducible con un estado aperiódico.

La siguiente proposición simplifica el estudio de las cadenas de Markov periódicas.

Proposición 8.2. *Consideremos una cadena de Markov irreducible, de período $d > 1$, con espacio de estados \mathbf{I} y matriz de transición $\mathbf{P} = (p_{ij})$. Entonces, existe una partición de \mathbf{I} en conjuntos disjuntos C_0, \dots, C_{d-1} , tales que $p_{ij} > 0$ solo si $i \in C_\alpha, j \in C_{[\alpha+1]_d}$, donde $[k]_d$ designa el resto de la división entera de k entre d .*

Demostración. Fijemos un estado $i \in \mathbf{I}$. Con ayuda de este estado, construimos los conjuntos C_0, \dots, C_{d-1} de la siguiente forma:

$$j \in C_\alpha \text{ si existe un natural } n \geq 0 \text{ tal que } p_{ij}^{nd+\alpha} > 0.$$

Como ejemplo, tenemos $i \in C_0$. Para verificar que la asignación es consistente (porque el natural n no tiene por que ser único), veamos que si $k \in C_\alpha \cap C_\beta$, entonces $\alpha = \beta$.

Como la cadena de Markov es irreducible, existe un natural m tal que $p_{ki}^m > 0$. Como $k \in C_\alpha$, existe n_α tal que $p_{ik}^{n_\alpha d + \alpha} > 0$, y por esto $p_{ii}^{n_\alpha d + \alpha + m} \geq$

$p_{ik}^{n\alpha+d+\alpha} p_{ki}^m > 0$. De aquí, por la definición de período, obtenemos que $[\alpha + m]_d = 0$. Razonando análogamente para β obtenemos que $[\beta + m]_d = 0$, de donde $\alpha = \beta$.

Sea ahora $j \in C_\alpha$. Veamos que si $p_{jk} > 0$ entonces $k \in C_{[\alpha+1]_d}$. Existe n con $p_{ij}^{nd+\alpha} > 0$. Entonces

$$p_{ik}^{nd+\alpha+1} \geq p_{ij}^{nd+\alpha} p_{jk} > 0,$$

de donde se obtiene que $k \in C_{[\alpha+1]_d}$, concluyendo la demostración. \square

Consideremos una cadena de Markov X_0, X_1, X_2, \dots irreducible y de período $d > 1$. De la proposición anterior obtenemos la siguiente conclusión: Si $\mathbf{P}(X_0 \in C_\alpha) = 1$, es decir, si la cadena de Markov parte de C_α (donde C_α es como en la proposición), entonces, la sucesión formada por X_0, X_d, X_{2d}, \dots es una cadena de Markov irreducible y aperiódica, con espacio de estados C_α , distribución inicial $(\pi_i)_{i \in C_\alpha}$, y matriz de transición $(p_{ij}^d)_{i \in C_\alpha, j \in C_\alpha}$.

8.3. Recurrencia

Consideremos una cadena de Markov X_0, X_1, X_2, \dots , con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π .

Introducimos la *probabilidad de la primera transición de i a j en n pasos*, mediante

$$f_{ij}^n = \mathbf{P}_i(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j).$$

Si $i = j$, f_{ii}^n es la *probabilidad del primer retorno a i en n pasos*. Designemos $\mathbf{A}_n = \{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}$ ($n = 1, 2, \dots$). Como $\mathbf{A}_n \subset \{X_n = j\}$, obtenemos que $f_{ij}^n = \mathbf{P}_i(\mathbf{A}_n) \leq \mathbf{P}_i(X_n = j) = p_{ij}^n$. Como los sucesos $\mathbf{A}_1, \mathbf{A}_2, \dots$ son incompatibles dos a dos, obtenemos que

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n$$

es la probabilidad la suma de estos sucesos, es decir, la *probabilidad de visitar j partiendo de i* . En el caso en el que $i = j$, f_{ii} es la *probabilidad de retornar a i* .

Definición 8.3. Decimos que un estado i es recurrente cuando $f_{ii} = 1$. De lo contrario, si $f_{ii} < 1$, decimos que i es un estado transitorio.

Consideremos para cada estado j la variable aleatoria

$$\tau_j = \inf\{n \geq 1: X_n = j\}. \quad (8.11)$$

Si el conjunto en (8.11) es vacío ponemos $\tau_j = \infty$. La variable aleatoria τ_j es el *tiempo del primer pasaje por j* cuando la distribución inicial es arbitraria, y es el *tiempo del primer retorno a j* cuando la cadena de Markov parte del estado j . Observemos que $\mathbf{A}_n = \{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\} = \{\tau_j = n\}$, de donde

$$f_{ij}^n = \mathbf{P}_i(\tau_j = n), \quad f_{ij} = \mathbf{P}_i(\tau_j < \infty).$$

En conclusión, un estado i es recurrente cuando $f_{ii} = \mathbf{P}_i(\tau_i < \infty) = 1$; transitorio, cuando $f_{ii} = \mathbf{P}_i(\tau_i < \infty) < 1$.

Ejemplo 8.5. Problemas de barrera en el paseo al azar simple.

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes, idénticamente distribuidas, con $\mathbf{P}(X_1 = 1) = p$, $\mathbf{P}(X_1 = -1) = q$, donde $0 < p < 1$, $p + q = 1$. Definimos las sumas $S_0 = 0$ y $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$). Tenemos entonces un *paseo al azar simple* (ver ejemplos 8.3 y 8.4). Consideramos un número entero $b > 0$ y queremos calcular f_{0b} , la probabilidad de visitar b partiendo del origen, que también se denomina *probabilidad de alcanzar la barrera b* .

Consideremos otro número entero $a < 0$, y la probabilidad

$$\alpha(i) = \mathbf{P}_i(\exists n \geq 0: S_n = b; S_m > a, m = 0, 1, \dots, n-1).$$

Esta cantidad es la *probabilidad de alcanzar la barrera de nivel b antes que la de nivel a* , y calcular esta probabilidad es resolver el *problema de dos barreras* para el paseo al azar simple.

Interpretemos también este modelo como un juego de apuestas sucesivas entre dos jugadores A y B, conocido como el *problema de la ruina del jugador*. El jugador A tiene un capital $-a$; el jugador B, un capital b . Si en la primer apuesta ocurre $\{X_1 = 1\}$ gana A, y recibe una unidad de B; si ocurre $\{X_1 = -1\}$ pierde A, y entrega una unidad a B; y así sucesivamente para $n = 2, 3, \dots$. Luego de la n -ésima apuesta el capital de A será $S_n - a$, y el capital de B será $b - S_n$. El capital total $S_n - a + b - S_n = b - a$,

es constante. La cantidad $\alpha(0)$ que queremos calcular es la probabilidad de que el jugador B pierda el juego (se arruine).

Aplicando la fórmula de la probabilidad total, tenemos

$$\begin{aligned}\alpha(i) &= \mathbf{P}_i(\exists n \geq 0: S_n = b; S_m > a, m = 0, 1, \dots, n-1 | S_1 = i+1)p \\ &\quad + \mathbf{P}_i(\exists n \geq 0: S_n = b; S_m > a, m = 0, 1, \dots, n-1 | S_1 = i-1)q \\ &= p\alpha(i+1) + q\alpha(i-1).\end{aligned}$$

Entonces la sucesión $\{\alpha(i)\}$ verifica, si $a < i < b$, una ecuación en diferencias finitas. Como $\alpha(a) = 0$ y $\alpha(b) = 1$ no es difícil resolver esta ecuación, obteniendo, si $p \neq q$, que

$$\alpha(i) = \frac{(q/p)^i - (q/p)^a}{(q/p)^b - (q/p)^a}, \quad i = a, a+1, \dots, b. \quad (8.12)$$

Observemos ahora, para calcular f_{0b} , que los sucesos $\mathbf{C}_a = \{\exists n \geq 0: S_n = b; S_m > a, m = 0, 1, \dots, n-1\}$ verifican $\mathbf{C}_a \subset \mathbf{C}_{a-1}$, y su suma, para todos los valores negativos de a , es el suceso $\{\exists n \geq 0: S_n = b\}$. Podemos entonces calcular f_{0b} tomando límite, si $a \rightarrow -\infty$ en (8.12). Supongamos primero que $p < q$. Tenemos

$$f_{0b} = \mathbf{P}_0(\exists n \geq 0: S_n = b) = \lim_{a \rightarrow -\infty} \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a} = \left(\frac{p}{q}\right)^b.$$

Consideremos ahora el caso $p \geq q$. Si $p = q = 1/2$, la solución de la ecuación en diferencias finitas es

$$\alpha(i) = \frac{i-a}{b-a} \quad i = a, a+1, \dots, b. \quad (8.13)$$

Por ésto, si $p \geq q$, tomando límite si $a \rightarrow -\infty$ (en la fórmula correspondiente, según el caso), tenemos

$$f_{0b} = \mathbf{P}_0(\exists n \geq 0: S_n = b) = \lim_{a \rightarrow -\infty} \alpha(0) = 1.$$

Como conclusión obtenemos que si $p < q$, la probabilidad de visitar $b > 0$ es $f_{0b} = (p/q)^b$; mientras que si $p \geq q$, tenemos $f_{0b} = 1$ y el estado b se visita (o la barrera b se alcanza) con probabilidad 1.

Para el estudio de la recurrencia en una cadena de Markov la siguiente fórmula será de utilidad:

$$p_{ij}^n = \sum_{m=1}^n f_{ij}^m p_{jj}^{n-m}. \quad (8.14)$$

Antes de su demostración veamos una aplicación, demostrando que:

$$f_{ij} > 0 \text{ si y solo si } i \rightarrow j. \quad (8.15)$$

En efecto, si $f_{ij} > 0$ existe n natural tal que $0 < f_{ij}^n \leq p_{ij}^n$, obteniendo $i \rightarrow j$. Por su parte, si $i \rightarrow j$, para algún natural n la suma en (8.14) es positiva, por lo que para algún natural m se tiene $f_{ij}^m > 0$, de donde $f_{ij} > 0$.

Veamos ahora la demostración de (8.14). Aplicando la fórmula de la probabilidad total, tenemos

$$\begin{aligned} p_{ij}^n &= \mathbf{P}_i(X_n = j) = \mathbf{P}_i\left(X_n = j, \bigcup_{m=1}^n \{\tau_j = m\}\right) \\ &= \sum_{m=1}^n \mathbf{P}_i(X_n = j, \tau_j = m) = \sum_{m=1}^n \mathbf{P}_i(X_n = j | \tau_j = m) \mathbf{P}_i(\tau_j = m) \\ &= \sum_{m=1}^n \mathbf{P}_i(\tau_j = m) \mathbf{P}_i(X_n = j | X_m = j) = \sum_{m=1}^n f_{ij}^m p_{jj}^{n-m}. \end{aligned}$$

La fórmula (8.14) nos permite demostrar los siguientes resultados.

Proposición 8.3. *Consideremos una cadena de Markov con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π .*

(a) *Criterio de recurrencia. Un estado i es recurrente si y solo si se verifica $\sum_{n=1}^{\infty} p_{ii}^n = \infty$.*

(b) *Consideremos dos estados i, j que se comunican, es decir $i \leftrightarrow j$. Si i es recurrente, entonces j es recurrente. En otras palabras, la recurrencia es una propiedad de clase.*

(c) *Si j es un estado recurrente y además $i \rightarrow j$, entonces $\sum_{n=1}^{\infty} p_{ij}^n = \infty$.*

(d) *Si j es un estado transitorio, entonces $\sum_{n=1}^{\infty} p_{ij}^n < \infty$ para todo estado i , y por ésto $\lim_{n \rightarrow \infty} p_{ij}^n = 0$.*

Observemos que la parte (d) en la proposición anterior da respuesta a las preguntas formuladas al inicio de esta sección, cuando el estado j es transitorio.

Demostración. Supongamos que $a_{ij} = \sum_{n=1}^{\infty} p_{ij}^n < \infty$. Aplicando la fór-

mula (8.14) y cambiando el orden en la suma, obtenemos

$$\begin{aligned} a_{ij} &= \sum_{n=1}^{\infty} p_{ij}^n = \sum_{n=1}^{\infty} \left(\sum_{m=1}^n f_{ij}^m p_{jj}^{n-m} \right) = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} f_{ij}^m p_{jj}^{n-m} \\ &= \sum_{m=1}^{\infty} f_{ij}^m \sum_{n=0}^{\infty} p_{jj}^n = f_{ij}(1 + a_{jj}), \end{aligned} \quad (8.16)$$

dado que $p_{jj}^0 = 1$.

Veamos ahora la demostración de (a). Supongamos primero que $a_{ii} = \sum_{n=1}^{\infty} p_{ii}^n < \infty$. Aplicando la fórmula (8.16) con $i = j$, resulta

$$f_{ii} = \frac{a_{ii}}{1 + a_{ii}} < 1,$$

de donde obtenemos que i es transitorio.

Supongamos ahora que $a_{ii} = \sum_{n=1}^{\infty} p_{ii}^n = \infty$, y veamos que i es recurrente. Para cada natural N tenemos

$$\sum_{n=1}^N p_{ii}^n = \sum_{n=1}^N \left(\sum_{m=1}^n f_{ii}^m p_{ii}^{n-m} \right) = \sum_{m=1}^N \sum_{n=m}^N f_{ii}^m p_{ii}^{n-m} \leq \sum_{m=1}^N f_{ii}^m \sum_{n=0}^N p_{ii}^n.$$

De aquí, obtenemos la acotación

$$f_{ii} \geq \sum_{m=1}^N f_{ii}^m \geq \frac{\sum_{n=1}^N p_{ii}^n}{1 + \sum_{n=1}^N p_{ii}^n} \rightarrow 1 \quad (N \rightarrow \infty).$$

En conclusión $f_{ii} = 1$, y el estado i es recurrente. Esto concluye la demostración de (a).

Veamos la demostración de (b). Como $i \leftrightarrow j$ existen r y s tales que $p_{ij}^r p_{ji}^s > 0$. Luego, para cada $n = 1, 2, \dots$, tenemos

$$p_{jj}^{n+r+s} \geq p_{ij}^r p_{ii}^n p_{ji}^s.$$

Como $\sum_{n=1}^{\infty} p_{ii}^n = \infty$, de la desigualdad anterior resulta que $\sum_{n=1}^{\infty} p_{jj}^n = \infty$, y aplicando (a) obtenemos que j es recurrente.

Veamos la demostración de (c). Como $i \rightarrow j$, según (8.15) $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^n > 0$, y existe n_0 tal que $f_{ij}^{n_0} > 0$. Para $n \geq n_0$, aplicando la fórmula (8.14), tenemos

$$p_{ij}^n \geq f_{ij}^{n_0} p_{jj}^{n-n_0},$$

de donde, sumando en los valores de n entre n_0 y N , obtenemos

$$\sum_{n=n_0}^N p_{ij}^n \geq f_{ij}^{n_0} \sum_{n=n_0}^N p_{jj}^{n-n_0}. \quad (8.17)$$

Como, en vista de (a), la serie a la derecha en (8.17) es divergente dado que j es recurrente, lo es también la serie a la izquierda, concluyendo la demostración de (c).

Demostración de (d). Como $a_{jj} = \sum_{n=1}^{\infty} p_{jj}^n < \infty$, ya que j es transitorio, aplicando la fórmula (8.16) obtenemos

$$a_{ij} = f_{ij}(1 + a_{jj}) \leq (1 + a_{jj}) < \infty,$$

concluyendo la demostración de (d), y de toda la proposición. \square

Ejemplo 8.6. Recurrencia en el paseo al azar simple.

Consideremos un paseo al azar simple, es decir una cadena de Markov con espacio de estados \mathbb{Z} , distribución inicial concentrada en el origen, y matriz de transición dada por

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p, \quad p_{ij} = 0, \quad \text{si } i \neq j \pm 1,$$

donde $0 < p < 1$ (ver ejemplo 8.4). Queremos estudiar la recurrencia de esta cadena de Markov. Como la recurrencia es una propiedad de clase y el paseo al azar simple es una cadena de Markov irreducible, estudiamos la recurrencia en un estado, por ejemplo el origen $i = 0$. Para ésto aplicamos el criterio de recurrencia (a) en la proposición 8.3. En primer lugar, tenemos $p_{00}^{2n+1} = 0$, porque la cadena de Markov considerada tiene período 2. Aplicando la fórmula (8.9) tenemos

$$p_{00}^{2n} = C_n^{2n} p^n q^n = \frac{(4p(1-p))^n}{\sqrt{n\pi}} (1 + \varepsilon_n),$$

con $\varepsilon_n \rightarrow 0$ ($n \rightarrow \infty$), donde aplicamos también la fórmula de Stirling (ver página 42). Entonces, si $p \neq 1/2$ tenemos $4p(1-p) < 1$, y en consecuencia $\sum_{n=0}^{\infty} p_{00}^{2n} < \infty$. En este caso el origen es un estado transitorio, y por lo tanto, todos los estados son transitorios. Si $p = 1/2$ tenemos $p_{00}^{2n} = (1/\sqrt{n\pi})(1 + \varepsilon_n)$, de donde se obtiene que $\sum_{n=0}^{\infty} p_{00}^{2n} = \infty$ y el origen, y en consecuencia todos los estados, son recurrentes.

En conclusión el paseo al azar simple es recurrente únicamente en el caso simétrico en el que $p = 1/2$, siendo transitorio si $p \neq 1/2$.

Veamos ahora que los estados recurrentes son esenciales.

Proposición 8.4. *Sean i, j dos estados en una cadena de Markov. Supongamos que i es recurrente y que $i \rightarrow j$. Entonces $j \rightarrow i$. En otras palabras, i es esencial. Además $f_{ji} = 1$.*

Demostración. Como i es recurrente tenemos $f_{ii} = 1$. Entonces

$$\begin{aligned}
 0 &= 1 - f_{ii} = \mathbf{P}_i(\tau_i = \infty) \\
 &= \sum_{n=1}^{\infty} \mathbf{P}_i(\tau_i = \infty, \tau_j = n) + \mathbf{P}_i(\tau_i = \infty, \tau_j = \infty) \\
 &\geq \sum_{n=1}^{\infty} \mathbf{P}_i(\tau_i = \infty, \tau_j = n) \\
 &= \sum_{n=1}^{\infty} \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n, X_{n+1} \neq i, \dots) \\
 &= \sum_{n=1}^{\infty} \mathbf{P}_i(X_{n+1} \neq i, \dots \mid X_n = j) \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n) \\
 &= \mathbf{P}_j(\tau_i = \infty) \sum_{n=1}^{\infty} \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n) \\
 &= (1 - f_{jj}) \sum_{n=1}^{\infty} \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n).
 \end{aligned}$$

Pero $\sum_{n=1}^{\infty} \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n) > 0$ ya que $i \rightarrow j$. Entonces $f_{ji} = 1$. Por (8.15) tenemos $j \rightarrow i$, concluyendo la demostración. \square

Como corolario de esta proposición obtenemos, que si j es no esencial, es transitorio. Aplicando entonces la parte (d) de la proposición 8.3, resulta que si j es no esencial, $\lim_{n \rightarrow \infty} p_{ij}^n = 0$ para todo estado i . Resta entonces estudiar el comportamiento asintótico de los estados recurrentes, lo que haremos en la próxima sección. Concluimos esta sección estudiando otro ejemplo de recurrencia.

Ejemplo 8.7. Paseos al azar simétricos en \mathbb{Z}^d y Teorema de Pólya.

Un *paseo al azar simétrico en \mathbb{Z}^d* es una cadena de Markov con espacio de estados \mathbb{Z}^d , el conjunto de las d -úplas de números enteros $i = (z_1, \dots, z_d)$, distribución inicial $\pi = (\pi_i)$ concentrada en el origen, es decir,

$\pi_0 = 1$, donde $0 = (0, \dots, 0)$, y $\pi_i = 0$ si $i \neq 0$, y matriz de transición definida como sigue. Distinguiamos los estados

$$e_1 = (1, 0, \dots, 0), e_2 = (0, 1, \dots, 0), \dots e_d = (0, 0, \dots, 1),$$

y definimos la matriz de transición $P = (p_{ij})$, mediante

$$p_{ij} = \begin{cases} 1/(2d), & \text{si } j - i = \pm e_k \ (k = 1, \dots, d), \\ 0, & \text{en otro caso.} \end{cases}$$

En otras palabras, dado un estado i son posibles únicamente las transiciones a los estados j tales que j e i difieren en una coordenada, y en esta coordenada difieren exactamente en una unidad, siendo todas las transiciones posibles equiprobables. Visto de otra manera, la cadena de Markov considerada, encontrándose en un estado i , elige con probabilidad $1/d$ que coordenada cambiar, y con probabilidad $1/2$ si aumentar o disminuir en una unidad esta coordenada (siendo ambos “sorteos” independientes). Es posible también formular este ejemplo mediante sumas $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$) de vectores aleatorios independientes X_1, X_2, \dots , cada uno de los cuales toma el valor $\pm e_k$ ($k = 1, \dots, d$) con probabilidad $1/(2d)$, y poner $S_0 = 0$.

Esta cadena de Markov es homogénea en el tiempo y homogénea en el espacio, como en el caso particular $d = 1$, en el que la cadena de Markov considerada aquí es el paseo al azar simple del ejemplo 8.4 con $p = q = 1/2$.

Queremos estudiar para cada $d \geq 2$ la recurrencia de esta cadena de Markov. No es difícil verificar que cualquier par de estados se comunica, resultando ser la cadena de Markov irreducible. Esto permite estudiar la recurrencia en un único estado, y elegimos el origen por simplicidad.

Calculemos la probabilidad de retornar al origen en n pasos, designada $p_{00}^n(d) = 0$ (donde, por conveniencia, se indica la dimensión d en la notación). Para retornar al origen tenemos que elegir, digamos, m_1 veces el vector e_1 y otras m_1 veces el vector $-e_1$; m_2 veces el vector e_2 y otras m_2 el vector $-e_2$; \dots ; m_d veces el vector e_d y m_d veces el vector $-e_d$. De aquí obtenemos que se retorna al origen únicamente en una cantidad par de pasos, y en consecuencia, $p_{00}^{2n+1}(d) = 0$. Además, si retornamos en $2n$ pasos, se verifica $m_1 + \dots + m_d = n$.

Como estas elecciones forman un conjunto de sucesos independientes, (dado que los vectores aleatorios X_1, X_2, \dots son independientes) tenemos una distribución multinomial con parámetros $(2n, p_1, \dots, p_{2d})$, donde $p_k =$

$1/(2d)$ ($k = 1, \dots, 2d$). Entonces, como los naturales m_k ($k = 1, \dots, d$) son arbitrarios, tenemos

$$p_{00}^{2n}(d) = \sum \frac{(2n)!}{(m_1!)^2 \cdots (m_d!)^2} \left(\frac{1}{2d}\right)^{2n}, \quad (8.18)$$

donde sumamos en todos los naturales m_1, \dots, m_d que verifican $m_1 + \cdots + m_d = n$.

Observemos que en el caso $d = 1$ se obtiene $m_1 = n$, la suma anterior tiene un único sumando, y tenemos

$$p_{00}^{2n}(1) = C_n^{2n} \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{n\pi}}(1 + \varepsilon_n), \quad (8.19)$$

con $\varepsilon_n \rightarrow 0$ ($n \rightarrow \infty$), como vimos en el ejemplo 8.6.

Consideremos el caso $d = 2$. Si $m_1 = m$ tenemos $m_2 = n - m$, y

$$\begin{aligned} p_{00}^{2n}(2) &= \sum_{m=0}^n \frac{(2n)!}{(m!)^2((n-m)!)^2} \left(\frac{1}{4}\right)^{2n} = \left(\frac{1}{4}\right)^{2n} C_n^{2n} \sum_{m=0}^n (C_m^n)^2 \\ &= \left(\frac{1}{4}\right)^{2n} (C_n^{2n})^2 = (p_{00}^{2n}(1))^2, \end{aligned} \quad (8.20)$$

donde en la ante última igualdad³ utilizamos la fórmula $\sum_{m=0}^n (C_m^n)^2 = C_n^{2n}$. Por ésto, en vista de (8.19), tenemos

$$p_{00}^{2n}(2) = \frac{1}{n\pi}(1 + \delta_n)$$

con $\delta_n \rightarrow 0$ ($n \rightarrow \infty$). Luego $\sum_{n=0}^{\infty} p_{00}^{2n}(2) = \infty$ y el origen, y en consecuencia todos los estados, son recurrentes en el caso $d = 2$.

Consideremos nuevamente el caso general con $d \geq 3$. Es inmediato verificar que la fórmula (8.18) se puede escribir, con $m_1 = m$, como

$$\begin{aligned} p_{00}^{2n}(d) &= \sum_{m=0}^n C_{2m}^{2n} \left(\frac{1}{d}\right)^{2m} \left(\frac{d-1}{d}\right)^{2n-2m} \frac{(2m)!}{(m!)^2} \left(\frac{1}{2}\right)^{2m} \\ &\quad \times \sum \frac{(2n-2m)!}{(m_2!)^2 \cdots (m_d!)^2} \left(\frac{1}{2(d-1)}\right)^{2n-2m}, \end{aligned}$$

³En el ejercicio 14 se propone un método alternativo para obtener la fórmula (8.20), mediante un argumento de independencia.

donde la segunda suma se realiza en los valores naturales de m_2, \dots, m_d que verifican $m_2 + \dots + m_d = n - m$. Entonces⁴

$$p_{00}^{2n}(d) = \sum_{m=0}^n C_{2m}^{2n} \left(\frac{1}{d}\right)^{2m} \left(\frac{d-1}{d}\right)^{2n-2m} p_{00}^{2m}(1) p_{00}^{2n-2m}(d-1). \quad (8.21)$$

Podemos ahora considerar el caso $d = 3$. Probaremos que

$$p_{00}^{2n}(3) \leq \frac{K}{n^{3/2}}, \quad (8.22)$$

resultando la serie $\sum_{n=1}^{\infty} p_{00}^{2n}(3)$ convergente, y en consecuencia transitorio el paseo al azar simétrico en \mathbb{Z}^3 . En efecto, considerando $d = 3$ en (8.21), tenemos

$$p_{00}^{2n}(3) = \sum_{m=0}^n C_{2m}^{2n} \left(\frac{1}{3}\right)^{2m} \left(\frac{2}{3}\right)^{2n-2m} p_{00}^{2m}(1) p_{00}^{2n-2m}(2).$$

En vista del estudio en los casos $d = 1$ y $d = 2$, existen constantes K_1 y K_2 tales que si $m = 0, 1, \dots, n$, valen las acotaciones

$$p_{00}^{2m}(1) \leq K_1 \frac{\sqrt{2n}}{2m+1}, \quad p_{00}^{2n-2m}(2) \leq \frac{K_2}{2n-2m+1}.$$

Utilizando estas acotaciones, tenemos

$$\begin{aligned} p_{00}^{2n}(3) &\leq \sum_{m=0}^n C_{2m}^{2n} \left(\frac{1}{3}\right)^{2m} \left(\frac{2}{3}\right)^{2n-2m} \frac{K_1 K_2 \sqrt{2n}}{(2m+1)(2n-2m+1)} \\ &= \frac{K_3 \sqrt{n}}{(2n+1)(2n+2)} \sum_{m=0}^n C_{2m+1}^{2n+2} \left(\frac{1}{3}\right)^{2m+1} \left(\frac{2}{3}\right)^{2n-2m+1} \\ &\leq \frac{K}{n^{3/2}} \sum_{m=0}^n C_{2m+1}^{2n+2} \left(\frac{1}{3}\right)^{2m+1} \left(\frac{2}{3}\right)^{2n-2m+1} \leq \frac{K}{n^{3/2}}, \end{aligned}$$

donde K es una constante, y utilizamos la acotación

$$\sum_{m=0}^n C_{2m+1}^{2n+2} \left(\frac{1}{3}\right)^{2m+1} \left(\frac{2}{3}\right)^{2n-2m+1} \leq \sum_{m=0}^{2n+2} C_m^{2n+2} \left(\frac{1}{3}\right)^m \left(\frac{2}{3}\right)^{2n+2-m} = 1.$$

⁴La fórmula (8.21) se puede obtener también directamente, calculando probabilidades condicionales.

Esto concluye la demostración de (8.22), y la demostración de que la cadena de Markov es transitoria, si $d = 3$. Análogamente se puede demostrar que $p_{00}^{2n}(4) \leq K/n^2$. Más en general, es posible establecer que $p_{00}^{2n}(d) \leq K/n^{d/2}$. Así se demuestra el siguiente resultado.

Teorema 8.1 (Pólya). *El paseo al azar simétrico en \mathbb{Z}^d es recurrente para $d = 1, 2$ y transitorio para $d \geq 3$.*

8.4. Probabilidades límites y distribuciones estacionarias

En esta sección estudiaremos el comportamiento asintótico de las probabilidades de transición para un estado recurrente, completando las respuestas a las preguntas (a) y (b) en la página 177.

Consideremos una cadena de Markov X_0, X_1, X_2, \dots , con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π . Definimos el *tiempo medio de retorno* a un estado i , mediante

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^n = \mathbf{E}_i \tau_i,$$

donde incluimos el caso $\mu_i = \infty$ si la serie anterior diverge. Decimos que un estado recurrente es *positivo* cuando $\mu_i < \infty$, decimos que un estado recurrente es *nulo*⁵, cuando $\mu_i = \infty$.

Teorema 8.2. *Consideremos un estado i recurrente y aperiódico. Entonces*

$$\lim_{n \rightarrow \infty} p_{ii}^n = 1/\mu_i,$$

donde $\mu_i = \sum_{n=1}^{\infty} n f_{ii}^n$ es el tiempo medio de retorno a i , y consideramos $1/\mu_i = 0$ cuando $\mu_i = \infty$.

Antes de comenzar la demostración verificamos que el período de un estado se puede calcular a través de las probabilidades del primer retorno f_{ii}^n , en lugar de las probabilidades de retorno p_{ii}^n (ver definición 8.2).

⁵El término *nulo* se debe a que las probabilidades límites para estos estados es nula (ver Teorema 8.3).

Designemos $d_f(i) = \text{m. c. d.}\{n: f_{ii}^n > 0\}$. Como tenemos la inclusión de conjuntos $\{n: f_{ii}^n > 0\} \subset \{n: p_{ii}^n > 0\}$, porque $f_{ii}^n \leq p_{ii}^n$, obtenemos que

$$d(i) \leq d_f(i). \quad (8.23)$$

El siguiente resultado establece la igualdad en la fórmula anterior.

Lema 8.1. *El período $d(i)$ de un estado i es el máximo común divisor del conjunto de los n naturales que verifican $f_{ii}^n > 0$. Es decir, $d(i) = d_f(i)$.*

Demostración. Sea n tal que $p_{ii}^n > 0$. Si $f_{ii}^n > 0$ entonces $d_f(i)$ divide a n . Si $f_{ii}^n = 0$, existe una sucesión de estados $i, i_1, \dots, i_{n-1}, i$ con $p_{ii_1} \cdots p_{i_{n-1}i} > 0$, que necesariamente contiene a i entre los estados intermedios. Entonces existen n_1 y n_2 que verifican $n_1 + n_2 = n$, y tales que $p_{ii}^{n_1} p_{ii}^{n_2} > 0$. Si $f_{ii}^{n_1} = 0$ ó $f_{ii}^{n_2} = 0$ se puede repetir el procedimiento anterior. Como n es finito, la aplicación del argumento anterior permite hallar naturales n_1, \dots, n_m tales que

$$n = n_1 + \cdots + n_m, \text{ y } f_{ii}^{n_1} \cdots f_{ii}^{n_m} > 0.$$

Entonces $d_f(i)$ divide a n , concluyendo que $d_f(i) \leq d(i)$. En vista de (8.23) concluimos la demostración. \square

Demostración del del teorema 8.2. Como el estado i es fijo abreviamos la notación, escribiendo $f_n = f_{ii}^n$, $p_n = p_{ii}^n$ para cada $n = 0, 1, \dots$; y $\mu = \mu_i$. Por ejemplo, la fórmula (8.14) con la notación introducida, es

$$p_n = \sum_{m=1}^n f_m p_{n-m}. \quad (8.24)$$

Tenemos que demostrar $p_n \rightarrow 1/\mu$ ($n \rightarrow \infty$), incluyendo el caso $p_n \rightarrow 0$ ($n \rightarrow \infty$) cuando $\mu = \infty$. Dividimos la demostración, por conveniencia, en 5 etapas.

Etapas 1. Sea $a_n = f_n + f_{n+1} + \cdots = \sum_{m=n}^{\infty} f_m$ ($n = 1, 2, \dots$). Para estas cantidades, cambiando el orden en la suma, tenemos

$$\mu = \sum_{n=1}^{\infty} n f_n = \sum_{n=1}^{\infty} \sum_{m=1}^n f_n = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} f_n = \sum_{m=1}^{\infty} a_m.$$

Además $f_n = a_n - a_{n+1}$ ($n = 1, 2, \dots$), resultando de (8.24), que

$$p_n = (a_1 - a_2)p_{n-1} + \cdots + (a_n - a_{n+1})p_0,$$

que reordenando, y teniendo en cuenta que $a_1 = \sum_{n=1}^{\infty} f_n = 1$, dado que i es recurrente, se escribe como

$$p_n a_1 + p_{n-1} a_2 + \cdots + p_0 a_{n+1} = p_{n-1} a_1 + p_{n-2} a_2 + \cdots + p_0 a_n.$$

Como además $p_0 = 1$, la aplicación sucesiva de la fórmula anterior nos permite obtener que para cada $n = 1, 2, \dots$, tiene lugar la identidad

$$p_n a_1 + p_{n-1} a_2 + \cdots + p_0 a_{n+1} = p_0 a_1 = 1, \quad (8.25)$$

concluyendo la primer etapa de la demostración.

Etapa 2. Sea $\alpha = \limsup_n p_n$. Consideremos una subsucesión $\{n_m\}$ tal que $\alpha = \lim_m p_{n_m}$. Eligiendo s arbitrario, que verifique $f_s > 0$, tenemos

$$\begin{aligned} \alpha &= \liminf_m p_{n_m} = \liminf_m \left\{ f_s p_{n_m-s} + \sum_{r=1, r \neq s}^{n_m} f_r p_{n_m-r} \right\} \\ &\leq f_s \liminf_m p_{n_m-s} + \limsup_m \left\{ \sum_{r=1, r \neq s}^{n_m} f_r p_{n_m-r} \right\} \\ &\leq f_s \liminf_m p_{n_m-s} + \left\{ \sum_{r=1, r \neq s}^{\infty} f_r \limsup_m p_{n_m-r} \right\} \\ &\leq f_s \liminf_m p_{n_m-s} + (1 - f_s) \alpha, \end{aligned}$$

donde hemos utilizado que $\sum_{r=1}^{\infty} f_r = 1$, y que $\limsup_m p_{n_m-r} \leq \alpha$, para cualquier $r = 1, 2, \dots$. De aquí obtenemos que

$$\alpha \leq \liminf_m p_{n_m-s}$$

y, según la definición de α , obtenemos que

$$\lim_m p_{n_m-s} = \alpha. \quad (8.26)$$

concluyendo la segunda etapa de la demostración.

Etapa 3. Veamos que existe un natural s' tal que

$$\lim_m p_{n_m-s} = \alpha, \text{ para todo } s \geq s'. \quad (8.27)$$

Sea $\{s_1, \dots, s_r\}$ un subconjunto finito del conjunto $\{n: f_n > 0\}$, elegido de tal forma que se verifique la propiedad $d(i) = \text{m. c. d.}\{s_1, \dots, s_r\} = 1$.

Sea $s' = \prod_{m=1}^r s_m$. Cualquier natural $s \geq s'$ se puede representar de la forma

$$s = s_1 t_1 + \cdots + s_r t_r,$$

donde t_1, \dots, t_r son números naturales. Como (8.26) vale para el primer natural $s = s_1$ del subconjunto finito considerado, tenemos $\lim_m p_{n_m - s_1} = \alpha$. Aplicando el razonamiento de la etapa 2 a la subsucesión $\{n_m - s_1\}_{m \geq 1}$ donde nuevamente tomamos $s = s_1$, obtenemos que $\lim_m p_{n_m - 2s_1} = \alpha$. Es claro entonces que si repetimos el procedimiento anterior un total de t_1 veces con $s = s_1$, obtenemos que $\lim_m p_{n_m - t_1 s_1} = \alpha$. Razonando análogamente para s_2, \dots, s_r , obtenemos (8.27).

Etapas 4. Consideremos un natural $s \geq s'$. De la fórmula (8.25) obtenemos

$$p_{n_m - s'} a_1 + p_{n_m - (s'+1)} a_2 + \cdots + p_{n_m - (s'+s)} a_{s+1} \leq 1.$$

Tomando límite cuando $m \rightarrow \infty$ en ambos miembros de la desigualdad anterior, resulta

$$\alpha(a_1 + \cdots + a_{s+1}) \leq 1,$$

y de aquí, en el caso en que $\mu = \sum_{n=1}^{\infty} a_n = \infty$, deducimos que $\alpha = \limsup_n p_n = 0$. Esto concluye la demostración en el caso en que el estado i es recurrente nulo (es decir, $\mu = \infty$). En el caso en que $\mu < \infty$, obtenemos la desigualdad

$$\alpha = \limsup_n p_n \leq 1/\mu. \quad (8.28)$$

Etapas 5. Consideremos el caso $\mu < \infty$. Sea ahora $\beta = \liminf_n p_n$, y consideremos una subsucesión $\{n_m\}$ tal que $\lim_m p_{n_m} = \beta$. Razonando en forma análoga a como lo hicimos en las etapas 2, 3 y 4 para el límite superior, obtenemos que existe un natural s'' tal que

$$\lim_m p_{n_m - s} = \beta, \text{ para todo } s \geq s''.$$

De la fórmula (8.25) deducimos que para todo $s \geq s''$ se verifica

$$\begin{aligned} 1 &\leq p_{n_m - s''} a_1 + p_{n_m - (s''+1)} a_2 + \cdots + p_0 a_{n_m - s''+1} \\ &\leq p_{n_m - s''} a_1 + p_{n_m - (s''+1)} a_2 + \cdots + p_{n_m - (s''+s)} a_{s+1} + \sum_{k=s+2}^{\infty} a_k. \end{aligned}$$

Tomando límite si $m \rightarrow \infty$ en ambos miembros de la desigualdad anterior, obtenemos

$$1 \leq \beta(a_1 + a_2 + \cdots + a_{s+1}) + \sum_{k=s+2}^{\infty} a_k$$

Tomando límite ahora cuando $s \rightarrow \infty$, obtenemos que

$$\beta = \liminf_n p_n \geq 1/\mu.$$

En vista de esta desigualdad y la desigualdad (8.28), deducimos que existe el límite $\lim_n p_n = 1/\mu$. Esto concluye la demostración del teorema. \square

Como primer consecuencia del teorema anterior, obtenemos que la clasificación de estados recurrentes en positivos y nulos es una propiedad de clase.

Proposición 8.5. *Consideremos dos estados aperiódicos i, j de una cadena de Markov que se comunican (es decir $i \leftrightarrow j$). Si i es recurrente positivo, j es recurrente positivo; si i es recurrente nulo, j es recurrente nulo.*

Demostración. Como $i \leftrightarrow j$ existen naturales r y s tales que $p_{ij}^r p_{ji}^s > 0$. Luego, para cada $n = 1, 2, \dots$, tenemos

$$p_{jj}^{n+s+r} \geq p_{ji}^s p_{ii}^n p_{ij}^r. \quad (8.29)$$

Supongamos que i es recurrente positivo. Entonces $\lim_n p_{ii}^n > 0$ y de (8.29) obtenemos que $\lim_n p_{jj}^n > 0$, concluyendo que j es recurrente positivo. Análogamente, si i es recurrente nulo, se tiene $\lim_n p_{ii}^n = 0$, y como para todo $n = 1, 2, \dots$ se cumple

$$p_{ii}^{n+s+r} \geq p_{ij}^r p_{jj}^n p_{ji}^s, \quad (8.30)$$

obtenemos que $\lim_n p_{jj}^n = 0$. En conclusión j es recurrente nulo. Esto concluye la demostración. \square

Estamos ahora en condiciones de dar respuesta completa a las preguntas (a) y (b) relativas al comportamiento asintótico de una cadena de Markov irreducible y aperiódica.

Teorema 8.3. *Consideremos una cadena de Markov irreducible y aperiódica con espacio de estados \mathbf{I} , matriz de transición \mathbf{P} , y distribución inicial π . Se dan únicamente las siguientes tres posibilidades:*

- (a) *La cadena de Markov es transitoria. Para cada par de estados i, j tenemos $\lim_{n \rightarrow \infty} p_{ij}^n = \lim_{n \rightarrow \infty} \pi_j^n = 0$. Mas aún, $\sum_{n=1}^{\infty} p_{ij}^n < \infty$.*

- (b) La cadena de Markov es recurrente nula. Para cada par de estados i, j tenemos $\lim_{n \rightarrow \infty} p_{ij}^n = \lim_{n \rightarrow \infty} \pi_j^n = 0$, pero ahora $\sum_{n=1}^{\infty} p_{ij}^n = \infty$.
- (c) La cadena de Markov es recurrente positiva. Para cada par de estados i, j tenemos $\lim_{n \rightarrow \infty} p_{ij}^n = \lim_{n \rightarrow \infty} \pi_j^n = 1/\mu_j > 0$, donde $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^n$ es el tiempo medio de retorno al estado j .

Demostración. Consideremos un estado arbitrario k . Tenemos tres posibilidades mutuamente excluyentes. O bien el estado es transitorio, o bien es recurrente positivo, o bien recurrente nulo. Como estas tres propiedades son de clase, y la cadena de Markov considerada es irreducible, resulta que tenemos necesariamente alguna de las tres alternativas en la tesis del teorema. Resta entonces demostrar que se verifican los límites en (a), (b) y (c).

Comencemos por (a). En la parte (d) de la proposición 8.3 obtuvimos la convergencia de la serie, y el resultado $\lim_n p_{ij}^n = 0$. Veamos que ocurre con la distribución en el instante n . Tomando límite si $n \rightarrow \infty$ en la fórmula (8.6), obtenemos

$$\lim_{n \rightarrow \infty} \pi_j^n = \lim_{n \rightarrow \infty} \sum_k \pi_k p_{kj}^n = \sum_k \pi_k \lim_{n \rightarrow \infty} p_{kj}^n = 0,$$

donde cambiamos el orden entre la suma y el límite, porque los coeficientes (π_k) suman 1, y las probabilidades p_{kj}^n están acotadas.

Consideremos ahora el caso (b). En la parte (c) de la proposición 8.3 obtuvimos la divergencia de la serie; en el teorema 8.2, el resultado $\lim_n p_{jj} = 0$. Si convenimos que $p_{jj}^m = 0$ si $m < 0$, en vista de la fórmula (8.14), podemos escribir

$$p_{ij}^n = \sum_{m=1}^{\infty} f_{ij}^m p_{jj}^{n-m}.$$

Tomando límite a ambos lados de esta igualdad si $n \rightarrow \infty$, obtenemos

$$\lim_n p_{ij}^n = \lim_n \sum_{m=1}^{\infty} f_{ij}^m p_{jj}^{n-m} = \sum_{m=1}^{\infty} f_{ij}^m \lim_n p_{jj}^{n-m}, \quad (8.31)$$

donde cambiamos ahora el orden entre la suma y el límite, porque los coeficientes (f_{ij}^m) suman 1 (como vimos en la proposición 8.4), y las probabilidades p_{jj}^n están acotadas. Luego $\lim_n p_{ij}^n = 0$. La verificación de $\lim_n \pi_j^n = 0$, es análoga a la realizada en (a).

Consideremos finalmente el caso (c). La fórmula (8.31) es válida, y como ahora $\lim_n p_{jj} = 1/\mu_j > 0$ (por el teorema 8.2), resulta que $\lim_n p_{ij}^n = 1/\mu_j$. Finalmente, la obtención del límite $\lim_n \pi_j^n = 1/\mu_j$ es también análoga al caso (a). \square

Cuando el espacio de estados \mathbf{I} es finito el comportamiento asintótico en una cadena de Markov irreducible y aperiódica es sencillo: siempre encontramos la alternativa (c).

Corolario 8.1. *Una cadena de Markov finita irreducible y aperiódica es recurrente positiva.*

Demostración. Si algún estado es transitorio, o recurrente nulo, lo son todos. Como siempre es posible intercambiar el orden entre un límite y una suma finita, tenemos

$$1 = \lim_n \sum_{j \in \mathbf{I}} p_{ij}^n = \sum_{j \in \mathbf{I}} \lim_n p_{ij}^n = 0,$$

obteniendo una contradicción. Entonces estamos en el caso (c) del teorema anterior, concluyendo la demostración. \square

Como conclusión del capítulo estudiamos como calcular los límites de las probabilidades de transición (que coinciden con los límites de las distribuciones de probabilidad), cuando consideramos una cadena de Markov irreducible y aperiódica con estados recurrentes positivos (es decir, el caso (c) en el teorema 8.3, y la pregunta (c) en la página 177).

Para este fin son de suma utilidad las *distribuciones estacionarias*, que definimos a continuación.

Definición 8.4. *Un vector $\nu = (\nu_i)_{i \in \mathbf{I}}$ con coordenadas no negativas, es una distribución estacionaria de una cadena de Markov con espacio de estados \mathbf{I} o de su matriz de transición $\mathbf{P} = (p_{ij})$, si se verifican la condiciones:*

$$(E1) \quad \sum_i \nu_i = 1,$$

$$(E2) \quad \nu_j = \sum_i \nu_i p_{ij} \text{ para todo estado } j.$$

En notación matricial, si designamos $\mathbf{1} = (1, 1, \dots)$, tenemos que ν es una distribución estacionaria si se verifican las condiciones:

$$(E1) \quad \nu \times \mathbf{1}^t = 1,$$

$$(E2) \quad \nu = \nu \times \mathbf{P}.$$

Supongamos ahora que la distribución inicial de una cadena de Markov es una distribución estacionaria. La distribución de probabilidad en tiempo n verifica (8.5), por lo que

$$\pi^n = \nu \times \mathbf{P}^n = (\nu \times \mathbf{P}) \times \mathbf{P}^{n-1} = \nu \times \mathbf{P}^{n-1} = \dots = \nu \times \mathbf{P} = \nu$$

y resulta constante, o *invariante* en el tiempo. Esta fórmula se escribe también, como

$$\nu_j = \sum_i \nu_i p_{ij}^n. \quad (8.32)$$

La existencia de una distribución estacionaria para una cierta cadena de Markov da información sobre su comportamiento asintótico, en particular, si la cadena es irreducible y aperiódica.

Teorema 8.4. *Consideremos una cadena de Markov irreducible y aperiódica con espacio de estados \mathbf{I} y matriz de transición \mathbf{P} . Entonces, la cadena de Markov es recurrente positiva si y solo si existe una distribución estacionaria $\nu = (\nu_i)$ de la matriz de transición \mathbf{P} .*

Además, en este caso, $\nu_i = 1/\mu_i > 0$ ($i \in \mathbf{I}$), donde $\mu_i < \infty$ es el tiempo medio de retorno a i , y por lo tanto, la distribución estacionaria es única.

Demostración. Supongamos que la cadena de Markov irreducible y aperiódica es recurrente positiva. Designemos $\nu_j = \lim_n p_{ij}^n = 1/\mu_j > 0$ ($j \in \mathbf{I}$) y veamos que $\nu = (\nu_j)_{j \in \mathbf{I}}$ es una distribución estacionaria. Primero, tenemos

$$\sum_j \nu_j = \sum_j \lim_n p_{ij}^n \leq \lim_n \inf \sum_j p_{ij}^n = 1,$$

donde aplicamos el lema de Fatou, y obtenemos $\sum_j \nu_j \leq 1$. Veamos que ν verifica (E2). Con un argumento similar, tenemos

$$\sum_i \nu_i p_{ij} = \sum_i \lim_n p_{ki}^n p_{ij} \leq \lim_n \inf \sum_i p_{ki}^n p_{ij} = \lim_n \inf p_{kj}^{n+1} = \nu_j.$$

Si para algún j_0 tenemos $\sum_i \nu_i p_{ij_0} < \nu_{j_0}$, entonces

$$\sum_j \nu_j > \sum_j \sum_i \nu_i p_{ij} = \sum_i \nu_i \sum_j p_{ij} = \sum_i \nu_i,$$

lo que es una contradicción. Entonces ν cumple (E2). Veamos que el vector ν verifica la condición (E1). Como vale (E2), tomando límite si $n \rightarrow \infty$ en la fórmula (8.32), obtenemos

$$\nu_j = \lim_n \sum_i \nu_i p_{ij}^n = \sum_i \nu_i \lim_n p_{ij}^n = \left(\sum_i \nu_i \right) \nu_j,$$

donde utilizamos que $\sum_j \nu_j \leq 1$ y que las probabilidades están acotadas. Como $\nu_j > 0$ deducimos que $\sum_i \nu_i = 1$, demostrando que ν es una distribución estacionaria.

Supongamos ahora que existe una distribución estacionaria $\nu = (\nu_i)$. Por el teorema 8.3 sabemos que existe $\lim_n p_{ij}^n$, y que no depende de i . Debemos verificar que $\lim_n p_{ij}^n = \nu_j > 0$. Tomando límite si $n \rightarrow \infty$ en (8.32) y aplicando la propiedad (E1), tenemos

$$\nu_j = \lim_{n \rightarrow \infty} \sum_i \nu_i p_{ij}^n = \sum_i \nu_i \lim_{n \rightarrow \infty} p_{ij}^n = \lim_{n \rightarrow \infty} p_{ij}^n.$$

Si $\lim_n p_{ij}^n = 0$ para algún estado j , lo mismo ocurre para todos los estados, contradiciendo $\sum_j \nu_j = 1$. Luego $\lim_n p_{ij}^n = \nu_j > 0$ para todo j , y la cadena de Markov es recurrente positiva. En vista de (c) en el teorema 8.3, $\nu_j = 1/\mu_j$ ($j \in \mathbf{I}$). Esto concluye la demostración. \square

Este teorema es una herramienta de suma utilidad práctica, dado que siempre es posible verificar la existencia de distribuciones estacionarias en cadenas de Markov finitas (se trata de resolver un sistema de ecuaciones lineales), y muchas veces esto también se puede hacer en cadenas de Markov con espacio de estados infinito, como muestra el siguiente ejemplo.

Ejemplo 8.8. Paseo al azar con barrera reflejante. Consideremos una cadena de Markov con espacio de estados $\mathbb{N} = \{0, 1, \dots\}$, el conjunto de los números naturales, y matriz de transición \mathbf{P} , dada por

$$p_{00} = 1 - r, \quad p_{01} = r, \quad p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p, \quad \text{para } i = 1, 2, \dots$$

con $0 < p < 1$ y $0 < r \leq 1$. Es sencillo de verificar que la cadena de Markov es irreducible y aperiódica, porque $r > 0$. Nos proponemos estudiar la recurrencia y calcular la distribución estacionaria, en los casos

que corresponda. El sistema de ecuaciones $\nu = \nu P$ es

$$\begin{aligned}(1-r)\nu_0 + (1-p)\nu_1 &= \nu_0 \\ r\nu_0 + (1-p)\nu_2 &= \nu_1 \\ p\nu_1 + (1-p)\nu_3 &= \nu_2 \\ &\dots \\ p\nu_{n-1} + (1-p)\nu_{n+1} &= \nu_n \\ &\dots\end{aligned}$$

Entonces, una vez fijado $\nu_0 = \alpha$ arbitrario, se verifica

$$\nu_n = \frac{r\alpha}{(1-p)} \left(\frac{p}{1-p} \right)^{n-1}, \quad n = 1, 2, \dots$$

Luego, para que exista una distribución estacionaria debe cumplirse $\nu_i \geq 0$ (esto se verifica si $\alpha \geq 0$) y $\sum_{n=0}^{\infty} \nu_n = 1$. Ésto, en particular, implica que la serie $\sum_{n=1}^{\infty} (p/(1-p))^n$ es convergente, lo que implica que $p < 1/2$. En este caso, si $\alpha = (1-2p)/(r+1-2p)$, ν es distribución estacionaria de la cadena de Markov considerada. Luego, en este caso, el paseo al azar con barrera reflejante es recurrente positivo.

Consideremos el caso $1/2 < p < 1$. Observemos que $f_{10} < 1$, dado que esta probabilidad coincide con la probabilidad del mismo suceso para el paseo al azar simple del ejemplo 8.5. Luego, según la proposición 8.4, la cadena de Markov es transitoria.

Consideremos ahora el caso $p = 1/2$. Aplicando la fórmula de la probabilidad total, tenemos

$$\begin{aligned}f_{11} &= \mathbf{P}_1(\exists n: X_n = 1) = \mathbf{P}_1(\exists n: X_n = 1 | X_1 = 2)/2 \\ &\quad + \mathbf{P}_1(\exists n: X_n = 1 | X_1 = 0)/2 = (f_{21} + f_{01})/2.\end{aligned}$$

Sabemos que $f_{21} = 1$, porque esta probabilidad coincide con la probabilidad del mismo suceso en el paseo al azar simple, calculada en el ejemplo 8.5. Para calcular f_{01} , tenemos

$$\begin{aligned}1 - f_{01} &= \mathbf{P}_0(X_m \neq 1, \forall m = 1, 2, \dots) \\ &= \lim_{n \rightarrow \infty} \mathbf{P}_0(X_m \neq 1, \forall m = 1, \dots, n) = \lim_{n \rightarrow \infty} (1-r)^n = 0,\end{aligned}$$

porque $r > 0$. Entonces $f_{01} = 1$, y de aquí resulta $f_{11} = 1$. Luego, la cadena de Markov es recurrente en el caso $p = 1/2$. Como no existe distribución estacionaria, la cadena de Markov es recurrente nula.

8.5. Ejercicios

1. Sea X_0, X_1, X_2, \dots una cadena de Markov homogénea en el tiempo. Demostrar que

$$\mathbf{P}(X_{n+k} = j | X_n = i) = \mathbf{P}(X_k = j | X_0 = i).$$

2. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, idénticamente distribuidas, con distribución lática (ver ejercicio 7, capítulo 6). Demostrar que la sucesión $S_0 = 0, S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$) es una cadena de Markov.

3. Se considera una cadena de Markov con matriz de transición

$$\mathbf{P} = \begin{bmatrix} 5/6 & 1/6 \\ 2/3 & 1/3 \end{bmatrix}$$

y con distribución inicial $\pi = (4/5, 1/5)$. Demostrar que π^n , la distribución de probabilidad en el instante n , verifica $\pi^n = \pi$, para cada $n = 1, 2, \dots$.

4. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes con distribución común de Poisson con parámetro $\lambda = 1$. Sea $S_0 = 0, S_n = X_1 + \dots + X_n$. (a) Observar que S_0, S_1, \dots es una cadena de Markov, y hallar su matriz de transición \mathbf{P} . (b) Hallar \mathbf{P}^n (Sugerencia: determinar la distribución de S_n .) (c) Determinar estados esenciales y no esenciales. (d) Hallar el límite cuando $n \rightarrow \infty$ de las probabilidades de transición $p_{0,i}^n$.

5. Consideremos una sucesión de variables aleatorias X_0, X_1, X_2, \dots que toma valores en el conjunto de los números enteros. (a) Decimos que la sucesión tiene *incrementos independientes*, cuando para cualquier elección de índices $0 \leq m_1 \leq n_1 \leq \dots \leq m_r \leq n_r$, las variables aleatorias $X_{m_1} - X_{n_1}, \dots, X_{m_r} - X_{n_r}$ son mutuamente independientes. Decimos que la sucesión tiene *incrementos estacionarios*, cuando la distribución de la variable aleatoria $X_{n+m} - X_n$ no depende de n . (a) Demostrar que si la sucesión dada tiene incrementos independientes, entonces verifica las condición (a) en la definición 8.1. (b) Demostrar que si la sucesión tiene incrementos independientes y estacionarios, entonces es una cadena de Markov homogénea en el tiempo.

6. Calcular las matrices de transición de orden n , en una cadena de Markov con matriz de transición

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

y determinar el período de sus estados.

7. Demostrar que una cadena de Markov irreducible, para la cual un elemento p_{jj} de la diagonal de la matriz de transición es positivo, no puede tener estados periódicos.

8. Demostrar que una cadena de Markov con matriz de transición $P = (p_{ij})$ y distribuciones en el instante n dadas por $\pi^n = (\pi_j^n)$, verifica

$$\pi_j^{m+n} = \sum_k \pi_k^m p_{kj}^n.$$

(Fórmula (8.4) de la página 175.)

9. *Procesos de ramificación.* Una población integrada inicialmente ($n = 0$) por un individuo ($X_0 = 1$) se reproduce de acuerdo a la siguiente regla: cada individuo que integre la población en el instante n tiene, en el instante $n + 1$, una cantidad $k = 0, 1, 2, \dots$ de hijos con probabilidad p_k , donde $p_k \geq 0$ y $\sum_{k=0}^{\infty} p_k = 1$, y él desaparece. (Alternativamente, se puede suponer que un objeto se parte en k pedazos con probabilidad p_k ($k = 1, 2, \dots$), o desaparece con probabilidad p_0 , y luego cada uno de esos pedazos se parte o desaparece con las mismas probabilidades, y así sucesivamente.) Sea X_n la cantidad de individuos de la población en tiempo n . (a) Demostrar que X_n es una cadena de Markov, identificar su espacio de estados \mathbf{I} , hallar la distribución inicial y la matriz de transición. (b) Suponiendo que $p_k > 0$ para todo $k = 0, 1, \dots$, determinar si hay estados no esenciales y estados absorbentes.

10. En el ejercicio anterior: (a) Sea $q = f_{10} = \mathbf{P}(X_n = 0 \text{ para algún } n)$ la probabilidad de extinción de la población. Encontrar la ecuación que cumple q . (b) Encontrar la probabilidad de extinción para la población si $p_0 = p_1 = 1/4$, $p_2 = 1/2$. (Este modelo se utiliza para determinar la probabilidad de que un cierto apellido desaparezca, contando los varones de una población.)

11. *Estados recurrentes.* Sea X_0, X_1, X_2, \dots una cadena de Markov con espacio de estados \mathbf{I} . Demostrar que si $i \rightarrow j$, entonces

$$\sum_{n=1}^{\infty} \mathbf{P}_i(X_1 \neq i, \dots, X_{n-1} \neq i, \tau_j = n) > 0,$$

donde $\tau_j = \inf\{n \geq 1: X_n = j\}$.

12. En el contexto del ejercicio anterior, se considera

$$g_{ij} = \mathbf{P}_i(X_n = j \text{ infinitas veces}) = \mathbf{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{X_n = j\}\right).$$

(a) Demostrar que $\mathbf{P}(|\{n: X_n = i\}| \geq m) = (f_{ii})^m$, donde $|A|$ es la cantidad de elementos de un conjunto A . (b) Concluir que si i es recurrente, entonces $g_{ii} = 1$, mientras que en otro caso, $g_{ii} = 0$.

13. En el contexto de los dos últimos ejercicios, demostrar que si j es recurrente y se cumple $i \leftrightarrow j$, entonces $g_{ij} = 1$. Sugerencia: demostrar la fórmula

$$1 = \sum_{n=1}^{\infty} \mathbf{P}_j(\tau_i = n) + \mathbf{P}_j(\tau_i = \infty) = f_{ji}g_{ij} + 1 - f_{ji}.$$

14. *Paseo al azar simple en \mathbb{Z}^2 .* Consideremos una cadena de Markov con espacio de estados \mathbb{Z}^2 , definida mediante $S_0 = (0, 0)$, $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$), donde X_1, X_2, \dots es una sucesión de vectores aleatorios independientes, cada una de los cuales toma uno de los cuatro valores $(\pm 1, 0)$, $(0, \pm 1)$, con probabilidad $1/4$. (a) Si $X_n = (A_n, B_n)$, demostrar que las variables aleatorias $A_n + B_n$ y $A_n - B_n$ son independientes, y determinar su distribución. (b) Observando que $\mathbf{P}(X_{2n} = 0) = P(A_1 + \dots + A_{2n} = 0, B_1 + \dots, B_{2n} = 0)$, calcular la probabilidad de retorno al origen en $2n$ pasos. (c) Estudiar la recurrencia para la cadena de Markov⁶.

15. Considérese el paseo al azar simétrico en \mathbb{Z}^4 (ver ejemplo 8.7, donde $d = 4$). Demostrar que existe una constante K positiva tal que

$$p_{00}^{2n}(d) \leq K/n^2.$$

Concluir que el paseo al azar es recurrente.

⁶Aquí se presenta una forma alternativa para estudiar la recurrencia del paseo al azar simétrico en \mathbb{Z}^2 visto en el ejemplo 8.7.

16. *Paseo al azar en un tetraedro.* Consideremos una partícula que pasea por los vértices de un tetraedro eligiendo, con equiprobabilidad, a que vértice ir: (i) entre los tres vértices accesibles desde el vértice en que se encuentra; (ii) entre los dos vértices accesibles, para no volver al vértice en que se encontraba. (a) Determinar en cual de los casos anteriores estamos en presencia de una cadena de Markov, determinar el conjunto de estados y la matriz de transición \mathbf{P} . (b) Discutir periodicidad, en el caso que corresponda. (c) Verificar que el vector de distribución uniforme discreta $\mathbf{u} = (1/n, \dots, 1/n)$ con n el número de estados, es invariante, es decir $\mathbf{u} = \mathbf{uP}$, y determinar el límite de las probabilidades de transición, así como el límite de las distribuciones de probabilidad.

17. Una matriz \mathbf{P} con entradas no negativas se dice *estocástica* (o de Markov) si la suma de los elementos de cada fila es la unidad. Una matriz \mathbf{P} se dice *doblemente estocástica* si además de ser estocástica, la suma de los elementos de cada columna también es la unidad. Si se trata de una matriz finita $n \times n$, sea $\mathbf{u} = (1/n, \dots, 1/n)$ el vector de probabilidades uniformes. Verificar, que si \mathbf{P} es doblemente estocástica, entonces $\mathbf{u} = \mathbf{uP}$. Encontrar el vector límite $\boldsymbol{\pi}$ de las probabilidades de transición, para una cadena de Markov finita, irreducible y aperiódica, con matriz doblemente estocástica.

18. *Paseo al azar simétrico en el plano con barreras reflejantes.* Consideremos un paseo al azar simétrico en \mathbb{Z}^2 (como en el ejercicio 14, o en la página 187). Restringimos el espacio de estados a un subconjunto de \mathbb{Z}^2 , que denominamos *región*. Decimos que la frontera de esta región es reflejante si cada vez que en un paseo al azar sin restricciones hubiese abandonado la región, éste es forzado a volver a la última posición. Probar que, si cada punto de la región es alcanzable desde cualquier otro, y si la región tiene una cantidad finita de puntos, entonces existe una distribución estacionaria. Hallarla.

Capítulo 9

Martingalas

Continuamos con el estudio de las sucesiones de variables aleatorias dependientes considerando las *martingalas*. Esta noción, proveniente de los juegos de azar secuenciales, utiliza para su definición la *esperanza condicional*, que estudiamos a continuación.

9.1. Esperanza condicional

Definición 9.1 (Esperanza condicional). *Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, una variable aleatoria X , y otra variable aleatoria Y con esperanza $\mathbf{E}Y$. La esperanza condicional de Y dada X , que designamos $\mathbf{E}(Y|X)$, es una variable aleatoria $g(X)$, donde la función $g(x)$ verifica la propiedad*

$$\mathbf{E} \mathbf{1}_{\{X \in I\}} Y = \mathbf{E} \mathbf{1}_{\{X \in I\}} g(X) \quad (9.1)$$

para todo intervalo $I = (a, b]$ de la recta real.

La propiedad (9.1) exige que las esperanzas de las variables aleatorias Y y $g(X)$ coincidan en los sucesos generados por X , es decir, en los sucesos de la forma $\{\omega: X(\omega) \in (a, b]\}$. En este sentido $\mathbf{E}(Y|X)$ es la función de X que mejor *aproxima*¹ a Y . En estadística matemática se dice que $\mathbf{E}(Y|X)$ es un *estimador* de la variable aleatoria Y , cuando observamos la variable aleatoria X .

¹Es posible demostrar cuando $\mathbf{E}Y^2 < \infty$, que $\mathbf{E}(Y|X)$ es la variable aleatoria que *minimiza* la distancia $[\mathbf{E}(Y - h(X))^2]^{1/2}$ entre las variables aleatorias $Y, h(X)$, donde $h(x)$ una función boreliana arbitraria.

Calcular la esperanza condicional es, según la definición, determinar una función $g(x)$. Veamos algunos ejemplos.

Ejemplo 9.1. Variables aleatorias degeneradas. Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, una variable aleatoria X , y otra variable aleatoria Y con esperanza $\mathbf{E}Y$.

(a) Si se verifica $X(\omega) = c$ ($\omega \in \Omega$), entonces

$$\mathbf{E}(Y | X) = \mathbf{E}Y, \quad (9.2)$$

es decir, la esperanza condicional es una constante. En efecto, la función constante $g(x) = \mathbf{E}Y$ verifica (9.1) porque, si $c \in I = (a, b]$ tenemos $\mathbf{1}_{\{X \in I\}} = 1$ ($\omega \in \Omega$), de donde $\mathbf{E} \mathbf{1}_{\{X \in I\}} Y = \mathbf{E}Y = \mathbf{E}(\mathbf{1}_{\{X \in I\}} \mathbf{E}Y)$, mientras que si $c \notin I$ ambos términos en (9.1) se anulan.

(b) Si se verifica $Y(\omega) = a$ ($\omega \in \Omega$) también vale (9.2). En este caso tenemos $\mathbf{E} \mathbf{1}_{\{X \in I\}} Y = \mathbf{P}(X \in I)a = \mathbf{E}(\mathbf{1}_{\{X \in I\}} a)$, para cualquier intervalo I .

Ejemplo 9.2. Variables independientes. Dadas dos variables aleatorias independientes X e Y tales que existe $\mathbf{E}Y$, también vale (9.2). En este caso, para cualquier intervalo $I = (a, b]$, tenemos

$$\mathbf{E} \mathbf{1}_{\{X \in I\}} Y = \mathbf{E} \mathbf{1}_{\{X \in I\}} \mathbf{E}Y = \mathbf{E}(\mathbf{1}_{\{X \in I\}} \mathbf{E}Y),$$

y se verifica (9.1).

Ejemplo 9.3. Dadas una variable aleatoria X y una función $h(x)$, tales que existe la esperanza $\mathbf{E}h(X)$, se tiene

$$\mathbf{E}(h(X) | X) = h(X), \quad (9.3)$$

dado que la función $h(x)$ verifica (9.1).

Determinemos ahora la función $g(x)$ en los casos en que el vector aleatorio (X, Y) tiene distribución discreta o absolutamente continua.

Esperanza condicional para variables aleatorias con distribución discreta

Consideremos un vector aleatorio (X, Y) con distribución discreta, que toma los valores (x_k, y_j) con probabilidades $p_{kj} = \mathbf{P}(X = x_k, Y = y_j)$ ($k, j = 1, 2, \dots$), y tal que $\sum_{k,j=1}^{\infty} |y_j| p_{kj} < \infty$.

Definimos la función $g(x)$ para los valores $x = x_k$ ($k = 1, 2, \dots$), mediante

$$g(x) = \sum_{j=1}^{\infty} y_j \mathbf{P}(Y = y_j | X = x), \quad (9.4)$$

y vemos que verifica (9.1): Dado $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}} Y &= \sum_{k: a < x_k \leq b} \sum_{j=1}^{\infty} y_j \mathbf{P}(Y = y_j | X = x_k) \mathbf{P}(X = x_k) \\ &= \sum_{k: a < x_k \leq b} g(x_k) \mathbf{P}(X = x_k) = \mathbf{E} \mathbf{1}_{\{X \in I\}} g(X). \end{aligned}$$

Tomando conjuntos de la forma $\{X = x_k\}$ es sencillo ver, que cualquier otra función $h(x)$ que verifique (9.1) verifica $h(x_k) = g(x_k)$ ($k = 1, 2, \dots$), resultando que $\mathbf{P}(g(X) = h(X)) = 1$. Cuando dos variables aleatorias X e Y verifican $\mathbf{P}(X = Y) = 1$, decimos que X e Y son iguales *casi seguramente*, y escribimos $X = Y$ c.s. Entonces, la esperanza condicional $g(X) = \mathbf{E}(Y | X)$ en el caso discreto es única, en el sentido anterior. (Si otra variable aleatoria $h(X)$ verifica (9.1), se cumple $h(X) = g(X)$ c.s.)

Esperanza condicional para variables aleatorias con distribución absolutamente continua

Consideremos un vector aleatorio (X, Y) con distribución absolutamente continua, densidad $p(x, y)$, y tal que $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |y| p(x, y) dx dy < \infty$. Dada $p_1(x) = \int_{-\infty}^{\infty} p(x, y) dy$, la densidad de la variable aleatoria X , la *densidad condicional* de Y dada X es $r(y|x) = p(x, y)/p_1(x)$, definida para los x reales que verifican $p_1(x) > 0$.

Definimos la función $g(x)$ para los x reales que verifican $p_1(x) > 0$, mediante

$$g(x) = \int_{-\infty}^{\infty} yr(y|x) dy, \quad (9.5)$$

y vemos que verifica (9.1): Dado $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}} Y &= \int_a^b \left(\int_{-\infty}^{\infty} yr(y|x) dy \right) p_1(x) dx = \int_a^b g(x) p_1(x) dx \\ &= \mathbf{E} \mathbf{1}_{\{X \in I\}} g(X). \end{aligned}$$

Al igual que en el caso discreto se puede ver que dada otra función $h(x)$ que verifique (9.1), tenemos $h(x) = g(x)$ para los x reales que verifican $p_1(x) > 0$, resultando que $\mathbf{P}(g(X) = h(X)) = 1$. Entonces, la esperanza condicional $g(X) = \mathbf{E}(Y | X)$ en el caso absolutamente continuo es única, en el mismo sentido que en el caso discreto.

Ejemplo 9.4. Probabilidad condicional. Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, dos sucesos \mathbf{A} y \mathbf{B} con $\mathbf{P}(\mathbf{A}) > 0$, y las variables aleatorias $X = \mathbf{1}_{\mathbf{A}}$, $Y = \mathbf{1}_{\mathbf{B}}$. El vector aleatorio (X, Y) tiene distribución discreta. Aplicando la fórmula (9.4), tenemos

$$g(x) = 1 \times \mathbf{P}(Y = 1 | X = x) + 0 \times \mathbf{P}(Y = 0 | X = x),$$

de donde $g(1) = \mathbf{P}(\mathbf{B} | \mathbf{A})$, $g(0) = \mathbf{P}(\mathbf{B} | \bar{\mathbf{A}})$. En conclusión,

$$\mathbf{E}(Y | X) = g(\mathbf{1}_{\mathbf{A}}) = \begin{cases} \mathbf{P}(\mathbf{B} | \mathbf{A}), & \text{si } \omega \in \mathbf{A}, \\ \mathbf{P}(\mathbf{B} | \bar{\mathbf{A}}), & \text{si } \omega \in \bar{\mathbf{A}}. \end{cases}$$

En otras palabras: si ocurre \mathbf{A} tenemos $\mathbf{E}(\mathbf{1}_{\mathbf{B}} | \mathbf{1}_{\mathbf{A}}) = \mathbf{P}(\mathbf{B} | \mathbf{A})$, si no, $\mathbf{E}(\mathbf{1}_{\mathbf{B}} | \mathbf{1}_{\bar{\mathbf{A}}}) = \mathbf{P}(\mathbf{B} | \bar{\mathbf{A}})$. En este sentido, la esperanza condicional es una generalización de la probabilidad condicional (1.6).

El ejemplo anterior motiva la siguiente definición: consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, una variable aleatoria X y un suceso \mathbf{B} . La *probabilidad condicional de \mathbf{B} dada X* , que designamos $\mathbf{P}(\mathbf{B} | X)$, es la esperanza condicional de la variable aleatoria $\mathbf{1}_{\mathbf{B}}$ dada X , es decir

$$\mathbf{P}(\mathbf{B} | X) = \mathbf{E}(\mathbf{1}_{\mathbf{B}} | X). \quad (9.6)$$

Ejemplo 9.5. Consideremos un vector aleatorio (X, Y) con densidad normal bidimensional $p(x, y)$ dada en (3.22). El vector aleatorio considerado tiene distribución absolutamente continua. La variable aleatoria X tiene distribución normal con parámetros (a_1, σ_1) (ver ejemplo 3.15), por lo que la densidad condicional de Y dada X está dada por

$$r(y | x) = \frac{1}{\sigma_2 \sqrt{2\pi(1 - \rho^2)}} \exp \left\{ \frac{-1}{2\sigma_2^2(1 - \rho^2)} (y - a_2 - \rho\sigma_2(x - a_1)/\sigma_1)^2 \right\}.$$

Esta función de y es la densidad de una variable aleatoria con distribución normal, con esperanza $a_2 + \rho\sigma_2(x - a_1)/\sigma_1$ y varianza $\sigma_2^2(1 - \rho^2)$. Aplicando (9.5) obtenemos que $g(x) = a_2 + \rho\sigma_2(x - a_1)/\sigma_1$. De aquí

$$\mathbf{E}(Y | X) = g(X) = a_2 + \rho\sigma_2(X - a_1)/\sigma_1.$$

Recordemos que si $\rho = 0$, las variables aleatorias X e Y son independientes (ejemplo 3.16). En este caso $\mathbf{E}(Y|X) = a_2 = \mathbf{E}Y$, como resulta también del ejemplo 9.2.

Ejemplo 9.6. Consideremos dos variables aleatorias independientes X, Y con densidades respectivas $p_1(x), p_2(y)$ y una función $h(x)$, tales que existe la esperanza $\mathbf{E}h(X + Y)$. Entonces $\mathbf{E}(h(X + Y)|X) = g(X)$, donde

$$g(x) = \mathbf{E}h(x + Y) = \int_{-\infty}^{\infty} h(x + y)p_2(y)dy.$$

En efecto, dado un intervalo $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E}\mathbf{1}_{\{X \in I\}}g(X) &= \int_a^b g(x)p_1(x)dx = \int_a^b \left(\int_{-\infty}^{\infty} h(x + y)p_2(y)dy \right) p_1(x)dx \\ &= \int_a^b \int_{-\infty}^{\infty} h(x + y)p_1(x)p_2(y)dxdy = \mathbf{E}\mathbf{1}_{\{X \in I\}}h(X + Y), \end{aligned}$$

de acuerdo a la fórmula (4.10), verificando (9.1).

La definición de martingala emplea la esperanza condicional de una variable aleatoria dado un vector aleatorio, que vemos a continuación.

Definición 9.2. Consideremos un vector aleatorio $F = (X_1, \dots, X_n)$, y una variable aleatoria Y con esperanza $\mathbf{E}Y$, definidos en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. La esperanza condicional de Y dado F , que designamos mediante $\mathbf{E}(Y|F)$, y también $\mathbf{E}(Y|X_1, \dots, X_n)$, es una variable aleatoria $g(F)$, donde la función $g(x)$ ($x \in \mathbb{R}^n$) verifica la propiedad

$$\mathbf{E}\mathbf{1}_{\{F \in I\}}Y = \mathbf{E}\mathbf{1}_{\{F \in I\}}g(F) \quad (9.7)$$

para todo $I = (a_1, b_1] \times \dots \times (a_n, b_n]$.

Si en la definición 9.2 tenemos $n = 1$, obtenemos la definición 9.1. En los casos en los que el vector aleatorio (X_1, \dots, X_n, Y) tiene distribución discreta o absolutamente continua, es sencillo obtener fórmulas para la función $g(x)$ ($x \in \mathbb{R}^n$) similares a (9.4) y (9.5), respectivamente.

Ejemplo 9.7. Sean X_1, \dots, X_N variables aleatorias independientes, idénticamente distribuidas, con esperanzas $\mathbf{E}X_1, \dots, \mathbf{E}X_N$. Veamos que para las sumas $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots, N$), se verifica

$$\mathbf{E}\left(\frac{S_n}{n} \mid S_{n+1}, \dots, S_N\right) = \frac{S_{n+1}}{n+1} \quad (n = 1, \dots, N-1). \quad (9.8)$$

En primer lugar, la variable aleatoria $S_{n+1}/(n+1)$ es función del vector aleatorio $F = (S_{n+1}, \dots, S_N)$. Resta entonces verificar (9.7). Considerando $I = I_{n+1} \times \dots \times I_N$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{F \in I\}} \frac{S_n}{n} &= \frac{1}{n} \sum_{k=1}^n \mathbf{E} \mathbf{1}_{\{F \in I\}} X_k = \mathbf{E} \mathbf{1}_{\{F \in I\}} X_1 \\ &= \frac{1}{n+1} \sum_{k=1}^{n+1} \mathbf{E} \mathbf{1}_{\{F \in I\}} X_k = \mathbf{E} \mathbf{1}_{\{F \in I\}} \frac{S_{n+1}}{n+1}, \end{aligned}$$

dado que $\mathbf{E} \mathbf{1}_{\{F \in I\}} X_k = \mathbf{E} \mathbf{1}_{\{F \in I\}} X_1$ para $k = 1, \dots, n+1$, concluyendo la demostración de (9.8).

Como conclusión de esta sección vemos que la esperanza condicional de la definición 9.1 existe y es única, cuando el vector aleatorio (X, Y) tiene distribución arbitraria.

Teorema 9.1. *Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, una variable aleatoria X , y otra variable aleatoria Y con esperanza $\mathbf{E}Y$. Existe una función boreliana $g(x)$ que verifica la propiedad (9.1). Además, esta función es única en el siguiente sentido: si existe otra función boreliana $h(x)$ que verifica (9.1), entonces $g(X) = h(X)$ c.s.*

Una consecuencia importante que resulta de la unicidad de la función $g(x)$ en el teorema anterior es la siguiente: para demostrar que una variable aleatoria $g(X)$ es la esperanza condicional de Y dada X , es suficiente verificar que $g(x)$ verifica la propiedad (9.1), es decir, que las esperanzas de Y y $g(X)$ coinciden en los sucesos generados por X . Un teorema análogo vale para la esperanza condicional dado un vector aleatorio, correspondiente a la definición 9.2².

Demostración. Sea $F(x, y)$ la función de distribución del vector aleatorio (X, Y) . Consideremos dos medidas μ y \mathbf{P}_X , definidas en los conjuntos borelianos B de la recta real, mediante

$$\begin{aligned} \mu(B) &= \int_{B \times \mathbb{R}} y dF(x, y) = \mathbf{E} \mathbf{1}_{\{X \in B\}} Y, \\ \mathbf{P}_X(B) &= \int_{B \times \mathbb{R}} dF(x, y) = \mathbf{E} \mathbf{1}_{\{X \in B\}} = \mathbf{P}(a < X \leq b). \end{aligned}$$

²El lector que no se encuentre familiarizado con la teoría de la medida puede restringirse a la consideración de variables aleatorias que tengan distribución discreta o absolutamente continua, y pasar a la sección siguiente.

Como existe $\mathbf{E}Y$, tenemos $\int_{\mathbb{R}^2} |y| dF(x, y) < \infty$, por lo que μ es una medida finita con signo, mientras que \mathbf{P}_X es una medida de probabilidad (ver sección 4.1). Además, si para algún B se verifica $\mathbf{P}_X(B) = \mathbf{E} \mathbf{1}_{\{X \in B\}} = 0$, se verifica también $\mu(B) = \mathbf{E} \mathbf{1}_{\{X \in B\}} Y = 0$. En otras palabras, la medida μ es absolutamente continua con respecto de la medida \mathbf{P}_X y, mediante la aplicación del teorema de Radon–Nykodym³, obtenemos que existe una función boreliana $g(x)$, llamada *derivada de Radon–Nikodym*, que verifica $\mu(B) = \int_B g(x) d\mathbf{P}_X$ para todo conjunto boreliano B de la recta real. En consecuencia, dado un intervalo $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}} Y &= \int_{I \times \mathbb{R}} y dF(x, y) = \mu(I) = \int_I g(x) d\mathbf{P}_X \\ &= \mathbf{E} \mathbf{1}_{\{X \in I\}} g(X), \end{aligned}$$

es decir, la función $g(x)$ verifica (9.1). (La última igualdad es la fórmula (4.3) aplicada a la función $\mathbf{1}_{\{x \in I\}} g(x)$.)

Si otra función boreliana $h(x)$ verifica (9.1), tenemos

$$\mathbf{E} \mathbf{1}_{\{X \in B\}} g(X) = \mathbf{E} \mathbf{1}_{\{X \in B\}} h(X)$$

para todo boreliano B de la recta real. En particular, si $C = \{x: g(x) > h(x)\}$, tenemos $\mathbf{P}_X(C) = 0$, de lo contrario, $\mathbf{E} \mathbf{1}_{\{X \in C\}} (g(X) - h(X)) > 0$, lo que es una contradicción; análogamente, si $D = \{x: g(x) < h(x)\}$, tenemos $\mathbf{P}_X(D) = 0$. De aquí obtenemos, que

$$\begin{aligned} \mathbf{P}(g(X) \neq f(X)) &= \mathbf{P}_X(g(x) \neq f(x)) \\ &= \mathbf{P}_X(C) + \mathbf{P}_X(D) = 0, \end{aligned}$$

obteniendo la unicidad de la función $g(x)$ (en el sentido del enunciado), y concluyendo la demostración. \square

9.2. Propiedades de la esperanza condicional

Consideramos, por simplicidad, las propiedades de la esperanza condicional dada una variable aleatoria correspondiente a la definición 9.1. Son válidas las mismas propiedades para la esperanza condicional dado un vector aleatorio de la definición 9.2.

³Ver Apéndice 3.5 en Borovkov [1].

Propiedad 1 (Linealidad). *Consideremos dos variables aleatorias Y_1, Y_2 con esperanzas respectivas $\mathbf{E}Y_1, \mathbf{E}Y_2$, una variable aleatoria X , y dos constantes a, b . Entonces*

$$\mathbf{E}(aY_1 + bY_2 | X) = a \mathbf{E}(Y_1 | X) + b \mathbf{E}(Y_2 | X).$$

Demostración. Verifiquemos que la función $h(x) = ag_1(x) + bg_2(x)$ verifica (9.1), donde $\mathbf{E}(Y_k | X) = g_k(X)$ ($k = 1, 2$). En efecto, dado $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}}(aY_1 + bY_2) &= a \mathbf{E} \mathbf{1}_{\{X \in I\}}Y_1 + b \mathbf{E} \mathbf{1}_{\{X \in I\}}Y_2 \\ &= a \mathbf{E} \mathbf{1}_{\{X \in I\}}g_1(X) + b \mathbf{E} \mathbf{1}_{\{X \in I\}}g_2(X) = \mathbf{E} \mathbf{1}_{\{X \in I\}}h(X). \end{aligned}$$

Entonces

$$\mathbf{E}(aY_1 + bY_2 | X) = h(X) = ag_1(X) + bg_2(X) = a \mathbf{E}(Y_1 | X) + b \mathbf{E}(Y_2 | X),$$

concluyendo la demostración. \square

Ejemplo 9.8. Sumas de variables aleatorias independientes. Consideremos la sucesión $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$), donde $\{X_n\}$ es una sucesión de variables aleatorias independientes con esperanzas respectivas $\{\mathbf{E}X_n\}$. Sea $S_0 = X_0 = 0$, y designemos $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$). Como existen las esperanzas $\{\mathbf{E}S_n\}$, podemos calcular

$$\begin{aligned} \mathbf{E}(S_{n+1} | F_n) &= \mathbf{E}(S_n + X_{n+1} | F_n) = \mathbf{E}(S_n | F_n) + \mathbf{E}(X_{n+1} | F_n) \\ &= S_n + \mathbf{E}X_{n+1}. \end{aligned}$$

donde utilizamos la linealidad de la esperanza condicional, la fórmula (9.3), y la independencia.

Propiedad 2 (Monotonía). *Consideremos una variable aleatoria X , y otra variable aleatoria Y con esperanza $\mathbf{E}Y$. Si tenemos $Y \geq 0$, se verifica $\mathbf{E}(Y | X) \geq 0$.*

Demostración. Si el vector aleatorio (X, Y) tiene distribución discreta, y toma los valores (x_k, y_j) ($k, j = 1, 2, \dots$), tenemos $y_j \geq 0$, por lo que $g(x) \geq 0$ (ver (9.4)) y $\mathbf{E}(Y | X) = g(X) \geq 0$.

Si el vector aleatorio (X, Y) tiene distribución absolutamente continua con densidad $p(x, y)$, la condición $Y \geq 0$ implica que $p(x, y) = 0$, si $y < 0$. Entonces $g(x) \geq 0$ (ver (9.5)) y $\mathbf{E}(Y | X) = g(X) \geq 0$.

En el caso general, si el lector considera posible analizar la demostración del teorema 9.1, observamos que la medida μ que allí se define es una medida positiva, y también lo es la derivada de Radon–Nikodym $g(x)$; concluyendo que $\mathbf{E}(Y|X) = g(X) \geq 0$ c.s. \square

Combinando las propiedades 1 y 2 obtenemos, que si X, Y_1, Y_2 son como en la propiedad 1, y además $Y_1 \leq Y_2$, entonces $\mathbf{E}(Y_1|X) \leq \mathbf{E}(Y_2|X)$.

Propiedad 3. *Consideremos una variable aleatoria X , y otra variable aleatoria Y con esperanza $\mathbf{E}Y$. Entonces*

$$\mathbf{E}\mathbf{E}(Y|X) = \mathbf{E}Y. \quad (9.9)$$

La igualdad (9.9) se denomina fórmula de la esperanza total.

Demostración. Sea $g(x)$ tal que $\mathbf{E}(Y|X) = g(X)$ y consideremos $I = (-\infty, \infty)$ en la propiedad (9.1). Como $\{\omega: X(\omega) \in I\} = \Omega$, tenemos $\mathbf{1}_{\{X \in I\}} = 1$ ($\omega \in \Omega$), y

$$\mathbf{E}Y = \mathbf{E}\mathbf{1}_{\{X \in I\}}Y = \mathbf{E}\mathbf{1}_{\{X \in I\}}g(X) = \mathbf{E}\mathbf{E}(Y|X),$$

lo que demuestra la propiedad. \square

Ejemplo 9.9. Consideremos sucesos $\mathbf{A}_1, \dots, \mathbf{A}_n$ incompatibles dos a dos, con probabilidades positivas, y tales que alguno de ellos ocurre. Sea \mathbf{B} un suceso arbitrario. A partir de los sucesos dados construimos las variables aleatorias $X = \sum_{k=1}^n k\mathbf{1}_{\mathbf{A}_k}$, $Y = \mathbf{1}_{\mathbf{B}}$.

El vector aleatorio (X, Y) tiene distribución discreta, y la función $g(x)$ en (9.4) vale

$$g(x) = \mathbf{P}(Y = 1 | X = x) \quad (x = 1, \dots, n).$$

Entonces, tenemos $g(k) = \mathbf{P}(\mathbf{B} | \mathbf{A}_k)$ ($k = 1, \dots, n$), por lo que $\mathbf{E}(Y|X) = g(X) = \mathbf{P}(\mathbf{B} | \mathbf{A}_k)$, si ocurre \mathbf{A}_k ($k = 1, \dots, n$). Aplicando la fórmula (9.9) de la esperanza total, obtenemos

$$\begin{aligned} \mathbf{P}(\mathbf{B}) &= \mathbf{E}Y = \mathbf{E}\mathbf{E}(Y|X) = \sum_{k=1}^n g(k) \mathbf{P}(X = k) \\ &= \sum_{k=1}^n \mathbf{P}(\mathbf{B} | \mathbf{A}_k) \mathbf{P}(\mathbf{A}_k), \end{aligned}$$

que es la fórmula de la probabilidad total (1.7). Este ejemplo muestra que la fórmula de la esperanza total (9.9) es una generalización de la fórmula de la probabilidad total (1.7).

Ejemplo 9.10. Sean N, X_1, \dots, X_K variables aleatorias independientes, tales que X_1, \dots, X_K son idénticamente distribuidas y existe $\mathbf{E} X_1 = a$. Supongamos además que la variable aleatoria N tiene distribución discreta, y toma una cantidad finita de valores $k = 1, \dots, K$. Queremos calcular la esperanza de la variable aleatoria

$$S = \sum_{k=1}^N X_k = X_1 + \dots + X_N, \quad (9.10)$$

que es la suma de una cantidad aleatoria de sumandos. Es sencillo ver que $\mathbf{E}|S| < \infty$. Veamos entonces que $\mathbf{E}(S|N) = aN$, verificando que la función $g(x) = ax$ verifica (9.1). En efecto, si $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{N \in I\}} S &= \sum_{n: a < n \leq b} \mathbf{E} \mathbf{1}_{\{N=n\}} (X_1 + \dots + X_n) = \sum_{n: a < n \leq b} na \mathbf{E} \mathbf{1}_{\{N=n\}} \\ &= \sum_{n: a < n \leq b} \mathbf{E} (\mathbf{1}_{\{N=n\}} aN) = \mathbf{E} (\mathbf{1}_{\{N \in I\}} aN), \end{aligned}$$

donde, en la segunda igualdad utilizamos la independencia y la equidistribución. Aplicando la fórmula (9.9) de la esperanza total, tenemos

$$\mathbf{E} S = \mathbf{E} \mathbf{E}(S|N) = a \mathbf{E} N = \mathbf{E} X_1 \mathbf{E} N.$$

Propiedad 4. Consideremos dos variables aleatorias X, Y y una función $h(x)$. Supongamos que existen las esperanzas $\mathbf{E} Y, \mathbf{E} (h(X)Y)$. Entonces,

$$\mathbf{E} (h(X)Y | X) = h(X) \mathbf{E}(Y | X). \quad (9.11)$$

Si ponemos $Y(\omega) = 1$ ($\omega \in \Omega$) en (9.11) obtenemos la fórmula (9.3) (ver ejemplo 9.1 (b)). Por ésto, la fórmula (9.11) es una generalización de la fórmula (9.3).

Demostración. Consideremos primero un vector aleatorio (X, Y) con distribución discreta que toma los valores (x_k, y_j) ($k, j = 1, 2, \dots$), la función $g(x)$ en (9.4), y veamos que la función $h(x)g(x)$ verifica (9.1). En efecto, si $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}} h(X)Y &= \sum_{k: a < x_k \leq b} h(x_k) \sum_{j=1}^{\infty} y_j \mathbf{P}(Y = y_j | X = x_k) \mathbf{P}(X = x_k) \\ &= \sum_{k: a < x_k \leq b} h(x_k) g(x_k) \mathbf{P}(X = x_k) = \mathbf{E} \mathbf{1}_{\{X \in I\}} h(X)g(X). \end{aligned}$$

Entonces $\mathbf{E}(h(X)Y | X) = h(X)g(X) = h(X) \mathbf{E}(Y | X)$, que es la fórmula (9.11).

Si el vector aleatorio (X, Y) tiene densidad $p(x, y)$, la función $h(x)g(x)$ verifica (9.1) con $g(x)$ definida en (9.5). En efecto, si $I = (a, b]$, tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X \in I\}} h(X)Y &= \int_a^b \left(\int_{-\infty}^{\infty} h(x)yr(y|x)dy \right) p_1(x)dx \\ &= \int_a^b h(x)g(x)p_1(x)dx = \mathbf{E} \mathbf{1}_{\{X \in I\}} h(X)g(X). \end{aligned}$$

y vale (9.11).

Consideremos finalmente el caso general, y sea $\mathbf{E}(Y | X) = g(X)$. Como la propiedad (9.1) se verifica para todo conjunto boreliano B de la recta real, la propiedad (9.11) se verifica si $h(x) = \sum_{k=1}^K c_k \mathbf{1}_{\{x \in B_k\}}$ es una función simple, donde c_1, \dots, c_K son reales arbitrarios y B_1, \dots, B_K son conjuntos borelianos arbitrarios. La demostración concluye, considerando una sucesión de funciones simples $\{h^n(x)\}$ que verifica $|h^n(x)| \leq |h(x)|$ y $h^n(X) \rightarrow h(X)$ ($n \rightarrow \infty$) *c.s.*, y aplicando el teorema de convergencia dominada. \square

Ejemplo 9.11. Productos de variables aleatorias independientes. Consideremos la sucesión $P_n = X_1 \times \dots \times X_n$ ($n = 1, 2, \dots$), donde $\{X_n\}$ es una sucesión de variables aleatorias independientes y no negativas con esperanzas respectivas $\{\mathbf{E} X_n\}$. Sea $P_0 = X_0 = 1$, y designemos $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$). Como existen las esperanzas $\{\mathbf{E} P_n\}$, podemos calcular

$$\mathbf{E}(P_{n+1} | F_n) = \mathbf{E}(P_n \times X_{n+1} | F_n) = P_n \mathbf{E}(X_{n+1} | F_n) = P_n \mathbf{E} X_{n+1}.$$

donde aplicamos la propiedad 4 y la independencia.

La última propiedad que consideramos es específica para la esperanza condicional dados vectores aleatorios.

Propiedad 5 (Telescópica). *Consideremos los vectores aleatorios $F_n = (X_1, \dots, X_n)$, $F_{n+m} = (X_1, \dots, X_n, \dots, X_{n+m})$, y una variable aleatoria Y con esperanza $\mathbf{E} Y$. Se verifican las igualdades:*

$$(a) \mathbf{E}(\mathbf{E}(Y | F_n) | F_{n+m}) = \mathbf{E}(Y | F_n),$$

$$(b) \mathbf{E}(\mathbf{E}(Y | F_{n+m}) | F_n) = \mathbf{E}(Y | F_n).$$

Demostración. (a) Sea $g(x)$ ($x \in \mathbb{R}^n$) tal que $\mathbf{E}(Y | F_n) = g(F_n)$. Considerando $g(F_n)$ función del vector aleatorio F_{n+m} y aplicando la fórmula (9.3), tenemos

$$\mathbf{E}(\mathbf{E}(Y | F_n) | F_{n+m}) = \mathbf{E}(g(F_n) | F_{n+m}) = g(F_n) = \mathbf{E}(Y | F_n),$$

lo que concluye la demostración de (a).

Demostremos (b). Consideremos la función $h(x)$ ($x \in \mathbb{R}^{n+m}$) tal que $\mathbf{E}(Y | F_{n+m}) = h(F_{n+m})$, y sea $k(z)$ ($z \in \mathbb{R}^n$) tal que $\mathbf{E}(h(F_{n+m}) | F_n) = k(F_n)$. Veamos que la función $k(z)$ verifica (9.7). En efecto, dado $I = (a_1, b_1] \times \cdots \times (a_n, b_n]$, tenemos $\{F_n \in I\} = \{F_{n+m} \in I \times \mathbb{R}^m\}$, y entonces

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{X_n \in I\}} k(F_n) &= \mathbf{E} \mathbf{1}_{\{X_n \in I\}} \mathbf{E}(h(F_{n+m}) | F_n) \\ &= \mathbf{E} \mathbf{E}(\mathbf{1}_{\{X_n \in I\}} h(F_{n+m}) | F_n) = \mathbf{E} \mathbf{E}(\mathbf{1}_{\{X_{n+m} \in I \times \mathbb{R}^m\}} h(F_{n+m}) | F_n) \\ &= \mathbf{E} \mathbf{1}_{\{X_{n+m} \in I \times \mathbb{R}^m\}} h(F_{n+m}) = \mathbf{E} \mathbf{1}_{\{X_{n+m} \in I \times \mathbb{R}^m\}} Y = \mathbf{E} \mathbf{1}_{\{X_n \in I\}} Y. \end{aligned}$$

Esto concluye la demostración de (b). \square

Consideramos finalmente la *desigualdad de Jensen*, relativa a intercambiar el orden de la esperanza condicional con la composición con una función convexa⁴.

Teorema 9.2. *Consideremos dos variables aleatorias X, Y , y una función real $\phi(y)$, convexa. Supongamos que existen las esperanzas $\mathbf{E}Y$, $\mathbf{E}\phi(Y)$. Se verifica*

$$\phi(\mathbf{E}(Y | X)) \leq \mathbf{E}(\phi(Y) | X). \quad (9.12)$$

La fórmula (9.12) se denomina desigualdad de Jensen (para la esperanza condicional).

Demostración. Consideremos un conjunto $\{y_1, y_2, \dots\}$ denso y numerable de puntos de la recta real. Para cada $n = 1, 2, \dots$, definimos

$$a_n = \phi'(y_n) = \lim_{y \rightarrow y_n^+} \frac{\phi(y) - \phi(y_n)}{y - y_n}, \quad b_n = \phi(y_n) - a_n y_n.$$

La constante a_n es la derivada por la derecha de la función $\phi(y)$ en el punto y_n (que siempre existe, dado que $\phi(y)$ es convexa). Como la función es convexa, obtenemos que

$$a_n y + b_n \leq \phi(y) \quad \text{para } y, n \text{ arbitrarios,} \quad (9.13)$$

⁴Decimos que una función real $\phi(x)$ es *convexa*, cuando $\phi(\lambda x + (1 - \lambda)y) \leq \lambda\phi(x) + (1 - \lambda)\phi(y)$ para x, y reales y $\lambda \in (0, 1)$, arbitrarios. Una función convexa es continua.

y como es continua, no es difícil verificar que

$$\phi(y) = \sup_n (a_n y + b_n). \quad (9.14)$$

Si sustituimos $y = Y$ en la desigualdad (9.13), aplicando la monotonía y la linealidad de la esperanza condicional, obtenemos

$$a_n \mathbf{E}(Y | X) + b_n \leq \mathbf{E}(\phi(Y) | X).$$

Para concluir la demostración tomamos supremo al variar $n = 1, 2, \dots$ en la desigualdad anterior, y nos referimos a (9.14). \square

Con el mismo esquema de demostración (tomando esperanza en vez de esperanza condicional en la fórmula (9.13)), se obtiene la *desigualdad de Jensen*: dadas una variable aleatoria Y y una función convexa $\phi(y)$, tales que existen las esperanzas $\mathbf{E}Y, \mathbf{E}\phi(Y)$, se verifica

$$\phi(\mathbf{E}Y) \leq \mathbf{E}\phi(Y). \quad (9.15)$$

Alternativamente, la desigualdad (9.15) se puede obtener, visto el ejemplo 9.1 (a), como un caso particular de la desigualdad (9.12) en el que $X(\omega) = 1$ ($\omega \in \Omega$).

9.3. Martingalas

A lo largo de esta sección suponemos dada una sucesión X_0, X_1, \dots de variables aleatorias, cuyas propiedades especificamos en cada situación cuando es necesario, y consideramos la sucesión de vectores aleatorios $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$).

Una sucesión Y_0, Y_1, \dots es *adaptada* a F_0, F_1, \dots cuando cada variable aleatoria Y_n es función del vector aleatorio F_n para cada $n = 0, 1, \dots$ (Es decir, para cada $n = 0, 1, \dots$ existe una función $f_n(x)$ ($x \in \mathbb{R}^{n+1}$), tal que $Y_n = f_n(F_n)$.) Decimos también que las variables aleatorias Y_0, \dots, Y_N son *adaptadas* a $\{F_0, \dots, F_N\}$ cuando cada Y_n es función de F_n ($n = 0, \dots, N$).

Definición 9.3 (Martingala, submartingala y supermartingala).

Decimos que una sucesión Y_0, Y_1, \dots de variables aleatorias con esperanzas respectivas $\mathbf{E}Y_0, \mathbf{E}Y_1, \dots$ y adaptada a $\{F_n\}$ es una martingala, cuando se verifica

$$\mathbf{E}(Y_{n+1} | F_n) = Y_n, \quad (9.16)$$

para $n = 0, 1, \dots$. La condición (9.16) se llama propiedad de martingala. Decimos que $\{Y_n\}$ es: una submartingala si vale $\mathbf{E}(Y_{n+1} | F_n) \geq Y_n$ en lugar de (9.16), una supermartingala si vale $\mathbf{E}(Y_{n+1} | F_n) \leq Y_n$ en lugar de (9.16).

Observemos que si $\{Y_n\}$ es una submartingala, entonces $\{-Y_n\}$ es una supermartingala, y que una martingala verifica $\mathbf{E} Y_n = \mathbf{E} Y_0$ ($n = 1, 2, \dots$), como resulta de tomar esperanza en (9.16) y aplicar la fórmula de la esperanza total (9.9). Es fácil ver, aplicando la propiedad telescópica 5, que la propiedad de martingala (9.16) es equivalente a

$$\mathbf{E}(Y_{n+m} | F_n) = Y_n, \quad (9.17)$$

para todo $n = 0, 1, \dots$ y todo $m = 1, 2, \dots$.

Ejemplo 9.12. Sumas y productos de variables aleatorias independientes. Consideremos las sumas de variables aleatorias independientes $\{S_n\}$ del ejemplo 9.8, que verifican

$$\mathbf{E}(S_{n+1} | F_n) = S_n + \mathbf{E} X_{n+1}.$$

Según la definición 9.3, la sucesión $\{S_n\}$ es una martingala cuando $\mathbf{E} X_n = 0$ ($n = 1, 2, \dots$); una submartingala cuando $\mathbf{E} X_n \geq 0$ ($n = 1, 2, \dots$); y una supermartingala cuando $\mathbf{E} X_n \leq 0$ ($n = 1, 2, \dots$).

A su vez, como los productos de variables aleatorias independientes y no negativas $\{P_n\}$ del ejemplo 9.11 verifican

$$\mathbf{E}(P_{n+1} | F_n) = P_n \mathbf{E} X_{n+1},$$

obtenemos que $\{P_n\}$ es una martingala cuando $\mathbf{E} X_n = 1$ ($n = 1, 2, \dots$), una submartingala cuando $\mathbf{E} X_n \geq 1$ ($n = 1, 2, \dots$), y una supermartingala cuando $\mathbf{E} X_n \leq 1$ ($n = 1, 2, \dots$).

Ejemplo 9.13. Martingalas y funciones convexas. Consideremos una sucesión Y_0, Y_1, \dots de variables aleatorias adaptadas a $\{F_n\}$, y una función real $\phi(y)$, tales que existen las esperanzas $\{\mathbf{E} \phi(Y_n)\}$.

(a) Si $\{Y_n\}$ es una martingala y la función $\phi(y)$ es convexa, la sucesión $\{\phi(Y_n)\}$ es una submartingala. En efecto, aplicando la desigualdad de Jensen (9.12), tenemos

$$\mathbf{E}(\phi(Y_{n+1}) | F_n) \geq \phi(\mathbf{E}(Y_{n+1} | F_n)) = \phi(Y_n),$$

y se verifica la definición de submartingala. Consideremos por ejemplo, la sucesión $\{S_n\}$ del ejemplo 9.8, y supongamos además que se verifica $\mathbf{E} X_n = 0$, $\mathbf{E} X_n^2 < \infty$ ($n = 1, 2, \dots$). Entonces, existe $\mathbf{E} S_n^2 = \mathbf{var} S_n$ (propiedad 5), y como la función $\phi(y) = y^2$ es convexa, obtenemos que $\{S_n^2\}$ es una submartingala.

(b) Si $\{Y_n\}$ es una submartingala y la función $\phi(y)$ es convexa y no decreciente, entonces la sucesión $\{\phi(Y_n)\}$ también es una submartingala. En efecto, aplicando nuevamente la desigualdad de Jensen (9.12) y la monotonía de $\phi(x)$, tenemos

$$\mathbf{E}(\phi(Y_{n+1}) | F_n) \geq \phi(\mathbf{E}(Y_{n+1} | F_n)) \geq \phi(Y_n),$$

y se verifica la definición de submartingala. Considerando, por ejemplo, la función $\phi(y) = \max(y - a, 0) = (y - a)^+$, que para cualquier a real es convexa y no decreciente, resulta que la sucesión $\{(Y_n - a)^+\}$ es una submartingala. (Es fácil ver que existe $\mathbf{E}(Y_n - a)^+$.)

El siguiente ejemplo relaciona las martingalas con los juegos de azar.

Ejemplo 9.14. Martingalas y apuestas. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, cada una de las cuales toma el valor 1 con probabilidad p ($0 < p < 1$), y el valor -1 con probabilidad $1 - p$.

Supongamos que un jugador que apuesta un monto b a la ocurrencia del suceso $\{X_n = 1\}$ ($n = 1, 2, \dots$) recibe $2b$ si acierta el resultado. En otras palabras, su capital aumenta en b si acierta y disminuye en b si no acierta.

Supongamos además que en el instante n , el jugador apuesta de acuerdo a una regla que tiene en cuenta los resultados anteriores⁵, es decir, el monto de su apuesta al suceso $\{X_{n+1} = 1\}$ es $b_n = b_n(X_1, \dots, X_n)$. Pongamos $X_0 = 0$. Si el jugador comienza a apostar con un capital inicial $Y_0 = y_0$ (constante), e Y_n representa su capital luego del n -ésimo resultado ($n = 1, 2, \dots$), se cumple la relación

$$Y_{n+1} = Y_n + X_{n+1} b_n(X_1, \dots, X_n) \quad (n = 0, 1, \dots),$$

de donde obtenemos, que

$$Y_n = y_0 + \sum_{k=0}^{n-1} (Y_{k+1} - Y_k) = y_0 + \sum_{k=0}^{n-1} X_{k+1} b_k(X_1, \dots, X_k),$$

⁵Éste parece ser el uso corriente de la palabra “martingala”.

que es una función de X_1, \dots, X_n . Como la variable aleatoria Y_{n+1} toma una cantidad finita de valores, su esperanza está dada por (4.5), y podemos calcular el *capital esperado* por el jugador para el turno $n + 1$, mediante

$$\begin{aligned}\mathbf{E}(Y_{n+1} | F_n) &= Y_n + b_n(X_1, \dots, X_n) \mathbf{E} X_{n+1} \\ &= Y_n + (p - q)b_n(X_1, \dots, X_n).\end{aligned}$$

donde $F_n = (X_0, \dots, X_n)$. En el caso $p = 1/2$ resulta que la sucesión Y_0, Y_1, \dots es una martingala, y decimos que el juego es *justo*; si $p > 1/2$ es una submartingla, y decimos que el juego es *favorable*; si $p < 1/2$ es una supermartingala, y decimos que el juego es *desfavorable*. Es interesante remarcar que ninguna sucesión de reglas $\{b_n\}$ permite, en un juego justo, aumentar el capital esperado.

Una resultado clave en la teoría de las martingalas es la posibilidad de sustituir los tiempos habituales $(n, n + 1, \dots)$ por una la clase tiempos aleatorios, que definimos a continuación.

Definición 9.4 (Tiempo de parada). *Una variable aleatoria τ que toma los valores $0, 1, \dots, \infty$ es un tiempo de parada⁶ con respecto de $\{F_n\}$, cuando el suceso $\{\tau = n\}$ se expresa a través de las variables aleatorias X_0, \dots, X_n , es decir, cuando la variable aleatoria $\mathbf{1}_{\{\tau=n\}}$ es una función de $F_n = (X_0, \dots, X_n)$ para cada $n = 0, 1, \dots$.*

Dado un natural n , la variable aleatoria $\tau(\omega) = n$ ($\omega \in \Omega$) verifica la definición anterior, por lo que los tiempos de parada son una generalización de los tiempos habituales. El ejemplo fundamental de tiempo de parada es el siguiente.

Ejemplo 9.15. Consideremos una sucesión de variables aleatorias $\{X_n\}$ y un intervalo arbitrario I de la recta real. La variable aleatoria

$$\tau = \inf\{n \geq 0: X_n \in I\},$$

con $\tau = \infty$ si el conjunto anterior es vacío, es un tiempo de parada, porque $\{\tau = n\} = \{X_0 \notin I, \dots, X_{n-1} \notin I, X_n \in I\}$.

Consideremos una sucesión Y_0, Y_1, \dots adaptada a $\{F_n\}$ y un tiempo de parada τ con respecto de $\{F_n\}$. Supongamos que $0 \leq \tau \leq N$. Introducimos

⁶Obsérvese que la variable aleatoria τ puede tomar el valor ∞ . Estrictamente hablando, se trata de una variable aleatoria *generalizada*.

la variable aleatoria

$$Y_\tau = \sum_{n=0}^N Y_n \mathbf{1}_{\{\tau=n\}},$$

que representa el valor de la sucesión $\{Y_n\}$ en el instante τ . Si designamos $F_\tau = (X_0, X_{\tau \wedge 1}, \dots, X_{\tau \wedge N})$, descomponiendo el conjunto Ω en los sucesos $\{\tau = k\}$ ($k = 0, \dots, N$), obtenemos que la variable aleatoria Y_τ es una función del vector aleatorio F_τ .

9.4. Teorema del muestreo opcional

En esta sección, al igual que en la anterior, suponemos dada una sucesión X_0, X_1, \dots de variables aleatorias, cuyas propiedades especificamos cuando es necesario; y consideramos la sucesión de vectores aleatorios $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$). El siguiente teorema explica como se sustituye un tiempo habitual por un tiempo de parada en la definición de martingala.

Teorema 9.3 (Teorema del muestreo opcional).

Consideremos una martingala Y_0, Y_1, \dots adaptada a $\{F_n\}$ y un tiempo de parada τ con respecto de $\{F_n\}$, que verifica $0 \leq \tau \leq N$ ($\omega \in \Omega$). Entonces

(a) $\mathbf{E}(Y_N | F_\tau) = Y_\tau,$

(b) $\mathbf{E} Y_N = \mathbf{E} Y_\tau = \mathbf{E} Y_0.$

Observación. El teorema del muestreo opcional 9.3 es válido también para submartingalas (supermartingalas), escribiendo \geq (\leq) en (a) y en (b), en vez de $=$.

Demostración. Como vimos que la variable aleatoria Y_τ es función del vector aleatorio F_τ , para demostrar (a) tenemos que verificar (9.7). Sea entonces $I = (a_0, b_0] \times \dots \times (a_n, b_n]$. Observando que el suceso $\{F_\tau \in I\} \cap \{\tau = k\}$ se expresa mediante las variables aleatorias X_0, \dots, X_k , tenemos

$$\begin{aligned} \mathbf{E} \mathbf{1}_{\{F_\tau \in I\}} Y_N &= \sum_{k=0}^N \mathbf{E} \mathbf{1}_{\{F_\tau \in I\} \cap \{\tau=k\}} Y_N \\ &= \sum_{k=0}^N \mathbf{E} (\mathbf{1}_{\{F_\tau \in I\} \cap \{\tau=k\}} \mathbf{E}(Y_N | F_k)) = \end{aligned}$$

$$= \sum_{k=0}^N \mathbf{E} \mathbf{1}_{\{F_\tau \in I\} \cap \{\tau=k\}} Y_k = \mathbf{E} \mathbf{1}_{\{F_\tau \in I\}} Y_\tau.$$

donde aplicamos la fórmula de la esperanza total (9.9), la propiedad (9.3), y la propiedad de martingala en la forma (9.17). Esto prueba entonces (a). La propiedad (b) se obtiene tomando esperanza en (a) y aplicando la fórmula de la esperanza total (9.9). \square

La extensión de este resultado para tiempos de parada generales, no acotados, exige hipótesis adicionales como muestra el siguiente ejemplo: consideremos las sumas de variables aleatorias independientes $\{S_n\}$ del ejemplo 9.8, donde además suponemos que $\mathbf{P}(X_n = 1) = \mathbf{P}(X_n = -1) = 1/2$ ($n = 1, 2, \dots$) de forma que $\{S_n\}$ es una martingala. Para el tiempo de parada

$$\tau = \inf\{n \geq 0: S_n = 1\},$$

se verifica $\mathbf{P}(\tau < \infty) = 1$ (como vimos en el ejemplo 8.6), pero tenemos

$$0 = \mathbf{E} S_0 \neq \mathbf{E} S_\tau = 1,$$

y la condición (b) en el teorema 9.3 del muestreo opcional no se verifica.

Ejemplo 9.16. Problemas de barrera en el paseo al azar simple II.

Consideremos el paseo al azar simple $\{S_n\}$ del ejemplo 8.5. Queremos resolver, aplicando ahora el teorema 9.3, el problema de dos barreras. Dados los enteros a, b , que verifican $a < 0 < b$, consideramos los tiempos de parada

$$\tau_a = \inf\{n \geq 0: S_n = a\}, \quad \tau_b = \inf\{n \geq 0: S_n = b\}.$$

En nuestro contexto, resolver el problema de dos barreras consiste en calcular $\mathbf{P}(\tau_b < \tau_a)$, la probabilidad de que el paseo al azar alcance la barrera de nivel b antes que la de nivel a .

Sea $\tau = \min(\tau_a, \tau_b) = \tau_a \wedge \tau_b$. Comencemos verificando que $\mathbf{P}(\tau < \infty) = 1$, es decir, que alguna de las dos barreras se alcanza. La variable aleatoria $\mu = (S_n + n)/2$ es la cantidad de éxitos en un esquema de Bernoulli (ver capítulo 2). Tenemos la igualdad de los sucesos

$$\begin{aligned} \mathbf{A}_n &= \{a < S_n < b\} = \left\{ \frac{a+n}{2} < \mu < \frac{b+n}{2} \right\} \\ &= \left\{ \frac{a-n(p-q)}{2\sqrt{npq}} < \frac{\mu-np}{\sqrt{npq}} < \frac{b-n(p-q)}{2\sqrt{npq}} \right\}. \end{aligned}$$

Además

$$\mathbf{P}(\tau = \infty) = \mathbf{P}(a < S_n < b, \text{ para todo } n) \leq \mathbf{P}(a < S_n < b) = \mathbf{P}(\mathbf{A}_n),$$

$$\delta_n = \Phi\left(\frac{b - n(p - q)}{2\sqrt{npq}}\right) - \Phi\left(\frac{a - n(p - q)}{2\sqrt{npq}}\right) \rightarrow 0 \quad (n \rightarrow \infty).$$

(Para obtener el último límite, hay que analizar por separado los casos $p = q$ y $p \neq q$.) Como la convergencia en el teorema límite integral de De Moivre–Laplace 2.2 es uniforme, obtenemos que $|\mathbf{P}(\mathbf{A}_n) - \delta_n| \rightarrow 0$ ($n \rightarrow \infty$), de donde resulta, que $\mathbf{P}(\tau = \infty) \leq \mathbf{P}(\mathbf{A}_n) \rightarrow 0$ ($n \rightarrow \infty$). En conclusión

$$\mathbf{P}(\tau < \infty) = \mathbf{P}(\tau_a < \tau_b) + \mathbf{P}(\tau_b < \tau_a) = 1. \quad (9.18)$$

Consideremos primero el caso $p \neq q$. Como $\mathbf{E}(q/p)^{X_n} = 1$ ($n = 1, 2, \dots$), la sucesión $\{Y_n = (q/p)^{S_n}\}$ es una martingala (ver ejemplo 9.12). Aplicando el teorema del muestreo opcional 9.3 con el tiempo de parada $\tau \wedge N = \min(\tau_a, \tau_b, N)$ (con $N \geq 1$ natural fijo), tenemos

$$1 = \mathbf{E} Y_1 = \mathbf{E} Y_{\tau \wedge N} = \mathbf{E}(q/p)^a \mathbf{1}_{\{\tau_a < \tau_b \wedge N\}} + \mathbf{E}(q/p)^b \mathbf{1}_{\{\tau_b < \tau_a \wedge N\}} + \mathbf{E}(q/p)^{S_N} \mathbf{1}_{\{\tau \geq N\}}. \quad (9.19)$$

La variable aleatoria $(q/p)^{S_N} \mathbf{1}_{\{\tau \geq N\}} \rightarrow 0$ ($N \rightarrow \infty$) *c.s.* y está uniformemente acotada, por lo que el último sumando en (9.19) tiende a cero si $N \rightarrow \infty$. Tomando límite en la igualdad anterior, obtenemos

$$(q/p)^a \mathbf{P}(\tau_a < \tau_b) + (q/p)^b \mathbf{P}(\tau_b < \tau_a) = 1. \quad (9.20)$$

Resolviendo el sistema formado por las ecuaciones lineales (9.18) y (9.20), obtenemos la probabilidad buscada:

$$\mathbf{P}(\tau_b < \tau_a) = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}, \quad (9.21)$$

que es el mismo resultado obtenido en (8.12), si $i = 0$.

Consideremos ahora el caso $p = q = 1/2$. Según vimos en el ejemplo (9.12), como $\mathbf{E} X_1 = 0$ la sucesión $\{S_n\}$ es una martingala, y obtenemos mediante la aplicación del teorema del muestreo opcional, en forma análoga a como lo hicimos en el caso $p \neq q$, que

$$a \mathbf{P}(\tau_a < \tau_b) + b \mathbf{P}(\tau_b < \tau_a) = 1. \quad (9.22)$$

Resolviendo ahora el sistema formado por (9.18) y (9.22), obtenemos

$$\mathbf{P}(\tau_b < \tau_a) = \frac{-a}{b-a},$$

que es el mismo resultado obtenido en (8.13), si $i = 0$.

La solución del problema de una barrera, es decir, el cálculo de la probabilidad $\mathbf{P}(\tau_b < \infty)$, se obtiene como en el ejemplo 8.5.

Otra aplicación del teorema del muestreo opcional es el siguiente resultado, que da información sobre la distribución del máximo de una submartingala.

Teorema 9.4. *Consideremos una submartingala $\{Y_n\}$. Para todo $\lambda > 0$ vale la desigualdad*

$$\lambda \mathbf{P} \left(\max_{0 \leq n \leq N} Y_n \geq \lambda \right) \leq \mathbf{E} Y_N^+. \quad (9.23)$$

La fórmula (9.23) se denomina desigualdad maximal de Doob.

Demostración. Consideremos el suceso

$$\mathbf{A} = \left\{ \max_{0 \leq n \leq N} Y_n \geq \lambda \right\}.$$

Aplicando el teorema del muestreo opcional a la submartingala $\{Y_n\}$ y al tiempo de parada $\tau = \inf\{n \geq 0: Y_n \geq \lambda\} \wedge N$, obtenemos

$$\mathbf{E} Y_N \geq \mathbf{E} Y_\tau = \mathbf{E} \mathbf{1}_{\mathbf{A}} Y_\tau + \mathbf{E} \mathbf{1}_{\mathbf{A}^c} Y_\tau \geq \lambda \mathbf{P}(\mathbf{A}) + \mathbf{E} \mathbf{1}_{\mathbf{A}^c} Y_N,$$

dado que si $\omega \in \mathbf{A}$, tenemos $Y_\tau \geq \lambda$. Entonces

$$\lambda \mathbf{P}(\mathbf{A}) \leq \mathbf{E} \mathbf{1}_{\mathbf{A}} Y_N \leq \mathbf{E} \mathbf{1}_{\mathbf{A}} Y_N^+ \leq \mathbf{E} Y_N^+,$$

lo que concluye la demostración. □

Como caso particular de la desigualdad maximal (9.23) obtenemos la *desigualdad de Kolmogorov* (9.24).

Corolario 9.1. *Sean X_0, \dots, X_N variables aleatorias independientes, que verifican $\mathbf{E} X_n = 0$, $\mathbf{E} X_n^2 < \infty$, para cada $n = 0, \dots, N$. Sea $S_n = X_0 + \dots + X_n$ ($n = 0, \dots, N$). Para todo $\lambda > 0$, se verifica la desigualdad*

$$\mathbf{P} \left(\max_{0 \leq n \leq N} |S_n| \geq \lambda \right) \leq \frac{1}{\lambda^2} \sum_{n=0}^N \mathbf{E} X_n^2. \quad (9.24)$$

Es interesante destacar que si $N = 0$ en (9.24) obtenemos la desigualdad de Chebishev (4.21), siendo entonces la desigualdad de Kolmogorov una generalización de la desigualdad de Chebishev.

Demostración. Como $\mathbf{E} X_n = 0$ ($n = 0, \dots, N$), aplicando la fórmula (4.19) obtenemos $\mathbf{E} S_N^2 = \text{var } S_N = \sum_{n=0}^N \mathbf{E} X_n^2$. Según vimos en el ejemplo 9.13 (a) la sucesión $\{S_n^2\}$ es una submartingala, y aplicando el teorema 9.4, obtenemos

$$\mathbf{P} \left(\max_{0 \leq n \leq N} |S_n| \geq \lambda \right) = \mathbf{P} \left(\max_{0 \leq n \leq N} S_n^2 \geq \lambda^2 \right) \leq \frac{1}{\lambda^2} \mathbf{E} S_N^2,$$

lo que concluye la demostración. \square

9.5. Convergencia de martingalas

Al igual que en la secciones anteriores, suponemos dada una sucesión X_0, X_1, \dots de variables aleatorias, cuyas propiedades especificamos cuando es necesario, y designamos $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$).

Teorema 9.5 (Convergencia de submartingalas de Doob).

Consideremos una submartingala $\{Y_n\}$ adaptada a $\{F_n\}$, que verifica la condición $\sup_n \mathbf{E} |Y_n| \leq C$, donde C es una constante. Entonces, existe el límite casi seguro $\lim_n Y_n = Y$ c.s., y se verifica $\mathbf{E} |Y| \leq C$.

La demostración de este teorema se basa en la *desigualdad de cruces*, también debida a Doob, que controla la oscilación de una submartingala.

Consideremos dos reales $a < b$, y definamos las variables aleatorias auxiliares:

$$\phi_0 = \begin{cases} 1, & \text{si } X_0 < a, \\ 0, & \text{en otro caso,} \end{cases}$$

$$\phi_n = \begin{cases} 1, & \text{si } \phi_{n-1} = 0 \text{ y } X_n < a, \\ 1, & \text{si } \phi_{n-1} = 1 \text{ y } X_n < b, \\ 0, & \text{en otro caso,} \end{cases}$$

para $n = 1, \dots, N$. En la figura 9.1 se indican los valores que toma esta sucesión para una trayectoria.

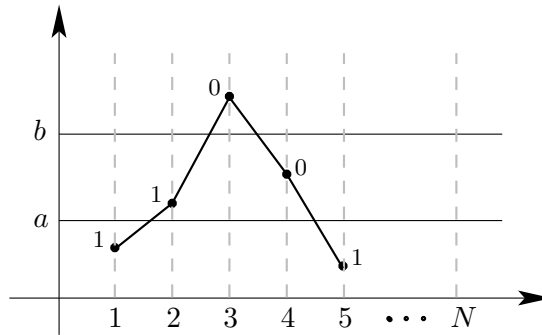


Figura 9.1: Se indica $\phi_1 = 1, \phi_2 = 1, \phi_3 = 0, \phi_4 = 0, \phi_5 = 1$.

Definimos la *cantidad de cruces* del intervalo $[a, b]$ hasta N , que designamos $U_N(a, b)$, mediante

$$U_N(a, b) = \sum_{n=0}^{N-1} \mathbf{1}_{\{\phi_n - \phi_{n+1} = 1\}} \quad (9.25)$$

que cuenta la cantidad de veces que cruzamos, en forma ascendente, el intervalo $[a, b]$, es decir, la cantidad de veces que la sucesión ϕ_0, \dots, ϕ_N pasa de 1 a 0.

Lema 9.1 (Desigualdad de cruces). *Consideremos las variables aleatorias Y_0, \dots, Y_N adaptadas a F_0, \dots, F_N , y dos reales $a < b$. Supongamos que se verifica la propiedad de submartingala*

$$\mathbf{E}(Y_{n+1} | F_n) \geq Y_n, \quad (n = 0, \dots, N - 1).$$

Entonces, se verifica

$$\mathbf{E} U_N(a, b) \leq \frac{1}{b - a} \mathbf{E}(Y_N - a)^+,$$

donde $U_N(a, b)$ está definido en (9.25).

Demostración. Consideremos las variables aleatorias $Z_n = (Y_n - a)^+$. Observemos que para cada $n = 0, \dots, N - 1$ se verifica $Z_{n+1}(\phi_n - \phi_{n+1}) \neq 0$ solamente cuando $\phi_n - \phi_{n+1} = 1$, porque si $\phi_n - \phi_{n+1} = -1$ tenemos

$\phi_n = 0$ y $\phi_{n+1} = 1$, por lo que $Z_{n+1} = 0$. Además, si $Z_{n+1}(\phi_n - \phi_{n+1}) \neq 0$ se tiene $Z_{n+1} \geq b - a$. Consideremos

$$\sum_{n=0}^{N-1} \phi_n (Z_{n+1} - Z_n) = \phi_N Z_N - \phi_0 Z_0 - \sum_{n=0}^{N-1} Z_{n+1} (\phi_{n+1} - \phi_n) \quad (9.26)$$

$$= \phi_N Z_N + \sum_{n=0}^{N-1} Z_{n+1} (\phi_n - \phi_{n+1})$$

$$\geq \sum_{n=0}^{N-1} Z_{n+1} \mathbf{1}_{\{\phi_n - \phi_{n+1} = 1\}}$$

$$\geq (b - a) \sum_{n=0}^{N-1} \mathbf{1}_{\{\phi_n - \phi_{n+1} = 1\}} = (b - a) U_N(a, b). \quad (9.27)$$

(La igualdad (9.26) es la fórmula de sumación por partes de Abel.) Las variables aleatorias Z_0, \dots, Z_N verifican la propiedad de submartingala, como vimos en el ejemplo 9.13 (b). Como además, cada variable aleatoria ϕ_n depende del vector F_n , aplicando la fórmula de la esperanza total (9.9) y la propiedad 4 para cada $n = 0, \dots, N - 1$, tenemos

$$\mathbf{E} \phi_n (Z_{n+1} - Z_n) = \mathbf{E} (\phi_n \mathbf{E}(Z_{n+1} - Z_n | F_n)) \leq \mathbf{E}(Z_{n+1} - Z_n).$$

Tomando entonces esperanza en los extremos de las desigualdades (9.26-9.27) obtenemos

$$(b - a) \mathbf{E} U_N(a, b) \leq \sum_{n=0}^{N-1} \mathbf{E}(Z_{n+1} - Z_n) = \mathbf{E} Z_N = \mathbf{E}(Y_N - a)^+,$$

lo que concluye la demostración. \square

Para continuar la demostración del teorema 9.5 utilizamos el siguiente resultado.

Lema 9.2. *Consideremos una sucesión Y_0, Y_1, \dots de variables aleatorias, tales que para cada par de números racionales $a < b$ y para cada N natural, se verifica $\mathbf{E} U_N(a, b) \leq K_{ab}$, donde K_{ab} es una constante independiente de N . Entonces, existe $Y = \lim_n Y_n$ c.s.*

Demostración. Dados $a < b$ racionales, consideremos el suceso

$$\begin{aligned} \mathbf{D}_{ab} &= \{\liminf_{n \rightarrow \infty} Y_n \leq a < b \leq \limsup_{n \rightarrow \infty} Y_n\} \\ &= \{U_n(a, b) \rightarrow \infty \ (n \rightarrow \infty)\}, \end{aligned}$$

que verifica

$$\{\omega: \lim_n Y_n(\omega) \text{ no existe}\} = \bigcup \{\mathbf{D}_{ab}: a < b \text{ racionales}\}. \quad (9.28)$$

Aplicando la desigualdad de Chebishev (4.21) para N y L arbitrarios, tenemos

$$\begin{aligned} \mathbf{P}(\mathbf{D}_{ab}) &= \mathbf{P}(U_n(a, b) \rightarrow \infty) \leq \mathbf{P}(U_N(a, b) \geq L) \\ &\leq \frac{1}{L} \mathbf{E} U_N(a, b) \leq \frac{K_{ab}}{L} \rightarrow 0 \quad (L \rightarrow \infty), \end{aligned}$$

Entonces se verifica $\mathbf{P}(\mathbf{D}_{ab}) = 0$, y como la suma de sucesos en (9.28) es numerable, tenemos $\mathbf{P}(\omega: \lim_n Y_n(\omega) \text{ no existe}) = 0$, concluyendo que existe $Y(\omega) = \lim_n Y_n(\omega)$ casi seguramente. \square

Demostración del teorema 9.5. De la aplicación de los lemas 9.1 y 9.2 obtenemos la existencia de $Y = \lim_n Y_n$ c.s. Aplicando el lema de Fatou obtenemos $\mathbf{E}|Y| = \mathbf{E} \lim_n |Y_n| \leq \liminf_n \mathbf{E}|Y_n| \leq C$, concluyendo la demostración del teorema. \square

Observemos que la convergencia en media de una submartingala exige hipótesis adicionales, como se ve a continuación.

Ejemplo 9.17. Consideremos los productos de variables aleatorias independientes $\{P_n\}$ del ejemplo 9.11, donde suponemos además que $\mathbf{P}(X_n = 0) = \mathbf{P}(X_n = 2) = 1/2$ ($n = 1, 2, \dots$). Como $\mathbf{E} X_n = 1$ la sucesión $\{P_n\}$ es una martingala, que verifica $\sup_n \mathbf{E}|P_n| = 1$. Aplicando el teorema 9.5, obtenemos que existe la variable aleatoria $P = \lim_n P_n$ c.s. En este caso particular podemos conocer la distribución de la variable aleatoria P . En efecto,

$$\begin{aligned} \mathbf{P}(P > 0) &= \mathbf{P}(P_k = 2 \text{ para todo } k = 1, 2, \dots) \\ &= \lim_{n \rightarrow \infty} \mathbf{P}(P_k = 2 \text{ para } k = 1, \dots, n) = \lim_{n \rightarrow \infty} 2^{-n} = 0. \end{aligned}$$

Es decir, la variable aleatoria P tiene distribución degenerada, con $\mathbf{P}(P = 0) = 1$. Por ésto, $\mathbf{E}|P_n - P| = \mathbf{E} P_n = 1$ y no hay convergencia en media.

Para estudiar la convergencia en media introducimos el siguiente concepto.

Definición 9.5 (Integrabilidad uniforme). *Decimos que una sucesión de variables aleatorias Y_0, Y_1, \dots es uniformemente integrable, si*

$$\lim_{H \rightarrow \infty} \sup_n \mathbf{E} \mathbf{1}_{\{|Y_n| > H\}} |Y_n| = 0.$$

Es inmediato verificar que si Y_0, Y_1, \dots es uniformemente integrable, entonces $\sup_n \mathbf{E} |Y_n| < \infty$. En efecto, dado $\varepsilon = 1$ existe H_0 , tal que se verifica $\mathbf{E} \mathbf{1}_{\{|Y_n| > H_0\}} |Y_n| \leq 1$ para todo n , por lo que

$$\sup_n \mathbf{E} |Y_n| \leq H_0 + \sup_n \mathbf{E} \mathbf{1}_{\{|Y_n| > H_0\}} |Y_n| \leq H_0 + 1.$$

(El recíproco de este resultado no es cierto.)

Ejemplo 9.18. Consideremos una sucesión $\{Y_n\}$ de variables aleatorias y otra variable aleatoria $Y \geq 0$ con esperanza $\mathbf{E}Y$, tales que se verifica $|Y_n| \leq Y$ ($n = 0, 1, \dots$). Entonces

$$\mathbf{E} \mathbf{1}_{\{|Y_n| > H\}} |Y_n| \leq \mathbf{E} \mathbf{1}_{\{Y > H\}} Y. \quad (9.29)$$

Como la acotación que obtenemos en (9.29) no depende del valor de n , y converge a cero si $H \rightarrow \infty$, resulta que la sucesión $\{Y_n\}$ es uniformemente integrable.

Concluimos el estudio de la convergencia de submartingalas con el siguiente resultado.

Teorema 9.6. *Una submartingala $\{Y_n\}$ uniformemente integrable converge casi seguramente y en media. Es decir*

(a) *Existe $Y = \lim_n Y_n$ c.s.*

(b) *Se verifica $\mathbf{E} |Y_n - Y| \rightarrow 0$, si $n \rightarrow \infty$.*

Demostración. Como la sucesión $\{Y_n\}$ es uniformemente integrable, existe una constante C tal que $\sup_n \mathbf{E} |Y_n| \leq C$. Aplicando el teorema 9.5, obtenemos que existe $Y = \lim_n Y_n$ c.s. y que se verifica $\mathbf{E} |Y| \leq C$.

Veamos (b). Sea $\varepsilon > 0$ arbitrario. Como existe $\mathbf{E}Y$ y la sucesión $\{Y_n\}$ es uniformemente integrable, existe H tal que

$$\sup_n \mathbf{E} \mathbf{1}_{\{|Y_n| > H\}} |Y_n| < \varepsilon/3, \quad \mathbf{E} \mathbf{1}_{\{Y > H\}} |Y| < \varepsilon/3. \quad (9.30)$$

Consideremos la función

$$h(y) = -H\mathbf{1}_{\{y < -H\}} + y\mathbf{1}_{\{-H \leq y \leq H\}} + H\mathbf{1}_{\{H < y\}},$$

que es continua, verifica $|h(y)| \leq H$, y además

$$|y - h(y)| \leq |y|\mathbf{1}_{\{|y| \geq H\}}. \quad (9.31)$$

Para todo n suficientemente grande tenemos

$$\mathbf{E} |h(Y_n) - h(Y)| \leq \varepsilon/3, \quad (9.32)$$

dado que $h(Y_n) - h(Y) \rightarrow 0$ ($n \rightarrow \infty$) *c.s.*, y $|h(Y_n) - h(Y)| \leq 2H$ ($n = 1, 2, \dots$). Podemos ahora concluir que, si n es suficientemente grande, tenemos

$$\mathbf{E} |Y_n - Y| \leq \mathbf{E} |Y_n - h(Y_n)| + \mathbf{E} |h(Y_n) - h(Y)| + \mathbf{E} |Y - h(Y)| < \varepsilon,$$

en vista de (9.31), (9.30), y (9.32). Hemos demostrado el teorema. \square

Observemos que en la demostración de la condición (b) no se utilizó la propiedad de submartingala.

9.6. Ley fuerte de los grandes números

Dedicamos esta sección a demostrar el siguiente resultado, obtenido por Kolmogorov.

Teorema 9.7 (Ley fuerte de los grandes números).

Consideremos una sucesión X_1, X_2, \dots de variables aleatorias independientes e idénticamente distribuidas, y tales que existe la esperanza $\mathbf{E} X_1 = a$. Entonces, tiene lugar la convergencia

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow a \quad \text{c.s.}$$

Demostración. Designemos $S_n = X_1 + \dots + X_n$ ($n = 1, 2, \dots$). La igualdad (9.8) del ejemplo 9.7 es la propiedad de martingala, con el tiempo invertido. En efecto, designando $Y_n = S_{N-n}/(N-n)$ ($n = 0, \dots, N-1$) y $F_n = (S_{N-n}, \dots, S_N)$, se verifica

$$\mathbf{E}(Y_{n+1} | F_n) = Y_n \quad (n = 0, \dots, N-2).$$

Estamos entonces en condiciones de aplicar el lema de cruces 9.1. Si $V_{N-1}(a, b)$ designa la cantidad de cruces ascendentes de Y_1, \dots, Y_{N-1} del intervalo $[a, b]$, tenemos $\mathbf{E} V_{N-1}(a, b) \leq \mathbf{E}(Y_{N-1} - a)^+ / (b - a) = \mathbf{E}(X_1 - a)^+ / (b - a)$. Consideremos ahora los cruces ascendentes $U_N(a, b)$ del vector aleatorio $(S_1, S_2/2, \dots, S_N/N)$ para el mismo intervalo $[a, b]$. Como este vector se obtiene a partir de Y_0, \dots, Y_{N-1} invirtiendo el tiempo, tenemos $U_N(a, b) \leq V_{N-1}(a, b) + 2$. Entonces,

$$\begin{aligned} \mathbf{E} U_N(a, b) &\leq 2 + \mathbf{E} V_{N-1}(a, b) \leq 2 + \frac{1}{b-a} \mathbf{E}(Y_N - a)^+ \\ &= 2 + \frac{1}{b-a} \mathbf{E}(X_1 - a)^+ \leq 2 + \frac{|a| + \mathbf{E}|X_1|}{b-a}. \end{aligned}$$

Estamos entonces en condiciones de aplicar el lema 9.2, de donde resulta que existe $\lim_n S_n/n$ casi seguramente. Para concluir la demostración aplicamos el teorema 5.6, de donde obtenemos que $S_n/n \xrightarrow{\mathbf{P}} a$. Como la convergencia casi segura implica la convergencia en probabilidad, y el límite en probabilidad es único, hemos concluido la demostración. \square

9.7. Ejercicios

En los ejercicios suponemos dada una sucesión X_0, X_1, \dots de variables aleatorias, cuyas propiedades especificamos cuando es necesario; y consideramos la sucesión de vectores aleatorios $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$).

1. Consideremos una variable aleatoria X con distribución discreta, que toma los valores x_1, x_2, \dots ; y otra variable aleatoria Y con esperanza $\mathbf{E} Y$. Demostrar que la función $g(x)$ definida para los valores $x = x_k$ ($k = 1, 2, \dots$) mediante

$$g(x) = \frac{1}{\mathbf{P}(X = x)} \mathbf{E} Y \mathbf{1}_{\{X=x\}}$$

verifica la propiedad (9.1).

2. *Varianza condicional.* Consideremos una variable aleatoria X , y otra variable aleatoria Y con varianza $\mathbf{var} Y$. Definimos la *varianza condicional* de Y dada X , que designamos $\mathbf{var}(Y | X)$, mediante

$$\mathbf{var}(Y | X) = \mathbf{E} ((Y - \mathbf{E}(Y | X))^2 | X).$$

(a) Demostrar la fórmula $\mathbf{var} Y = \mathbf{var} \mathbf{E}(Y | X) + \mathbf{E} \mathbf{var}(Y | X)$. (b) Consideremos la suma de una cantidad aleatoria de sumandos (9.10) del ejemplo 9.10, donde suponemos además que existe $\mathbf{var} X_1$. Demostrar que $\mathbf{var} S = a^2 \mathbf{var} N + \mathbf{E} N \mathbf{var} X_1$.

3. Sean X, Y variables aleatorias independientes, con esperanza nula, cada una de las cuales tiene distribución normal. (a) Demostrar que

$$\mathbf{E}((X - Y)^2 | X^2 + Y^2) = \mathbf{E}((X - Y)^2 | X^2, Y^2) = X^2 + Y^2.$$

(b) ¿Es válido el mismo resultado cuando las variables X e Y son independientes, simétricas, y tienen densidad?

4. Consideremos un vector aleatorio $F = (X_1, \dots, X_n)$, y una variable aleatoria Y con esperanza $\mathbf{E} Y$. Escribir fórmulas análogas a (9.4) y (9.5), en los casos en los que el vector aleatorio (X_1, \dots, X_n, Y) tiene distribución discreta o absolutamente continua.

5. La esperanza condicional permite definir cadenas de Markov en espacios de estados no necesariamente numerables. Consideremos una sucesión $\{X_n\}$ de variables aleatorias, que verifica

$$\mathbf{P}(X_{n+1} \in I | X_n, \dots, X_0) = \mathbf{P}(X_{n+1} \in I | X_n), \quad (9.33)$$

para todo $n = 1, 2, \dots$, y todo intervalo $I = [a, b]$. Demostrar que si las variables aleatorias $\{X_n\}$ toman valores en un conjunto \mathbf{I} finito o numerable, la definición (9.33) es equivalente a la definición 8.1 (a).

6. Consideremos variables aleatorias X, Y , tales que existe $\mathbf{E} Y$. Demostrar la desigualdad $\mathbf{E} |\mathbf{E}(Y | X)| \leq \mathbf{E} |Y|$, de dos formas: (a) sin utilizar la desigualdad de Jensen; (b) utilizándola.

7. (a) Demostrar la desigualdad de Jensen (9.15). (b) Obtener, a partir de la desigualdad de Jensen, la *desigualdad de Lyapunov*: Dada una variable aleatoria X con momento absoluto $\beta_r = \mathbf{E} |X|^r$ finito, se verifica $\beta_s^{1/s} \leq \beta_r^{1/r}$, si $0 < s < r$.

8. *Función característica de una variable de Poisson compuesta*. Consideremos una sucesión N, X_1, X_2, \dots de variables aleatorias independientes, tales que X_1, X_2, \dots son idénticamente distribuidas, la variable aleatoria N tiene distribución de Poisson con parámetro $\lambda > 0$. Demostrar

que la función característica de la variable aleatoria $S = \sum_{k=1}^N X_k = X_1 + \cdots + X_N$, está dada por

$$f(t) = \mathbf{E} e^{itS} = \exp \{ \lambda \mathbf{E}(e^{itX_1} - 1) \}.$$

9. *Definición de martingala.* Verificar que las propiedades (9.16) y (9.17) son equivalentes, para los índices m, n considerados. Demostrar que si $\{Y_n\}$ es martingala, se verifica $\mathbf{E} Y_n = \mathbf{E} Y_0$ ($n = 0, 1, \dots$); que si $\{Y_n\}$ es submartingala, se verifica $\mathbf{E} Y_n \leq \mathbf{E} Y_{n+1}$ ($n = 0, 1, \dots$); que si $\{Y_n\}$ es supermartingala, se verifica $\mathbf{E} Y_n \geq \mathbf{E} Y_{n+1}$ ($n = 0, 1, \dots$).

10. Consideremos una sucesión X_0, X_1, \dots de variables aleatorias que verifican $\mathbf{E} X_n = 0$ ($n = 0, 1, \dots$), y tales que existe $\mathbf{E} \exp X_n$ ($n = 0, 1, \dots$).
 (a) Demostrar que la sucesión $\{\exp(X_0 + \cdots + X_n)\}$ es una submartingala.
 (b) Encontrar constantes a_n tales que $\{\exp(X_0 + \cdots + X_n - a_n)\}$ sea una martingala.

11. *Descomposición de Doob.* Sea Y_0, Y_1, \dots una submartingala adaptada a $\{F_n\}$. (a) Demostrar que

$$M_n = Y_0 + \sum_{k=1}^n (Y_k - \mathbf{E}(Y_k | F_{k-1})) \quad (n = 1, 2, \dots)$$

es una martingala, y que la sucesión $A_n = Y_n - M_n$ ($n = 1, 2, \dots$) verifica $0 \leq A_1 \leq A_2 \leq \dots$, siendo la variable A_n función del vector F_{n-1} ($n = 2, 3, \dots$). La sucesión $\{A_n\}$ se llama *compensador* de $\{Y_n\}$.

(b) Calcular el compensador de $\{S_n^2\}$, donde $S_0 = 0$, $S_n = X_1 + \cdots + X_n$ ($n = 1, 2, \dots$), cuando X_1, X_2, \dots son variables aleatorias independientes, con esperanzas nulas, y tales que existe $\mathbf{E} X_n^2$, para todo $n = 1, 2, \dots$.

12. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, con densidad $p(x)$. Consideremos una función $h(x)$ que verifica

$$\int_{-\infty}^{\infty} h(x+y)p(y)dy = h(x) \quad \text{para todo } x \text{ real.} \quad (9.34)$$

Definamos $S_0 = X_0 = x$ y $S_n = x + X_1 + \cdots + X_n$ ($n = 1, 2, \dots$) Demostrar (suponiendo que existen $\{\mathbf{E} h(S_n)\}$) que $\{h(S_n)\}$ es una martingala. La identidad (9.34) se llama *ecuación de Wiener-Hopf*.

13. Sean τ y σ tiempos de parada con respecto de $\{F_n\}$; N un natural positivo. Determinar si son tiempos de parada las variables aleatorias siguientes: (a) $\tau + N$; (b) $\tau - N$; (c) $\max(\tau, \sigma)$; (d) $\min(\tau, \sigma)$; (e) $\tau + \sigma$.

14. Demostrar la siguiente variante del teorema del muestreo opcional: Sea X_0, X_1, \dots una martingala adaptada a $\{F_n\}$. Sean σ y τ tiempos de parada tales que se verifica $0 \leq \sigma \leq \tau \leq N$, con N un natural fijo. Se verifica: (a) $\mathbf{E}(X_\tau | F_\sigma) = X_\sigma$; (b) $\mathbf{E} X_0 = \mathbf{E} X_\sigma = \mathbf{E} X_\tau = \mathbf{E} X_N$.

15. *Desigualdades maximales.* Sea $\{Y_n\}$ una submartingala. Para todo $\lambda > 0$ valen la desigualdades:

$$(a) \lambda \mathbf{P} \left(\min_{0 \leq n \leq N} Y_n \leq -\lambda \right) \leq \mathbf{E} Y_N^+ - \mathbf{E} Y_0.$$

$$(b) \lambda \mathbf{P} \left(\max_{0 \leq n \leq N} |Y_n| \geq \lambda \right) \leq 3 \max_{0 \leq n \leq N} \mathbf{E} |Y_n|.$$

16. *Desigualdad maximal para supermartingalas.* Demostrar que si $\{Y_n\}$ es una supermartingala, vale la desigualdad (b) en el ejercicio anterior.

17. (a) Demostrar que si Y_0, Y_1, \dots es una supermartingala no negativa (es decir, $Y_n \geq 0$ ($n = 0, 1, \dots$)), entonces existe su límite casi seguro. (b) ¿Existe el límite en media?

18. Consideremos una variable aleatoria Z con esperanza $\mathbf{E} Z$ y definamos $Y_n = \mathbf{E}(Z | F_n)$ ($n = 0, 1, \dots$). Demostrar que $\{Y_n\}$ es una martingala, que converge casi seguramente, y en media.

19. *Integrabilidad Uniforme.* Consideremos una sucesión Y_0, Y_1, \dots de variables aleatorias.

(a) Se sabe que si la sucesión $\{Y_n\}$ es uniformemente integrable, entonces $\sup_n \mathbf{E} |Y_n| < \infty$. ¿El recíproco es cierto?

(b) Supongamos que $Y_n = Y$ ($n = 0, 1, \dots$). Verificar que $\{Y_n\}$ es uniformemente integrable si y solo si existe $\mathbf{E} Y$.

(c) Sabemos que si existe una variable aleatoria no negativa Y , con $\mathbf{E} Y < \infty$, y tal que $|Y_n| \leq Y$ para todo n , la sucesión $\{Y_n\}$ es uniformemente integrable. ¿El recíproco es cierto?

(d) Supongamos que $\sup_n \mathbf{E} |Y_n|^{1+\delta} \leq L$, donde $\delta > 0$ y L son constantes. Demostrar que la sucesión $\{Y_n\}$ es uniformemente integrable. ¿Es cierto el recíproco de esta proposición?

(e) *Criterio de La Vallée Pousin*: Si existe una función $\Phi: [0, \infty) \rightarrow [0, \infty)$ monótona creciente, tal que $\lim_{x \rightarrow \infty} \Phi(x)/x = \infty$, y $\sup_n \mathbf{E} \Phi(|Y_n|) < \infty$, entonces, la sucesión $\{Y_n\}$ es uniformemente integrable. (El recíproco de esta proposición es cierto⁷: dada $\{Y_n\}$ uniformemente integrable, existe una tal función Φ .)

20. Consideremos el espacio de sucesos elementales $\Omega = (0, 1]$, la σ -álgebra de Borel \mathcal{B} , junto con \mathbf{P} la medida de Lebesgue en Ω . Sea Z una variable aleatoria con esperanza matemática definida en este espacio de probabilidad. Para cada $n = 0, 1, \dots$ introducimos la variable aleatoria

$$X_n(\omega) = \sum_{k=0}^{2^n-1} k \mathbf{1}_{\{k2^{-n} < \omega \leq (k+1)2^{-n}\}},$$

y los vectores $F_n = (X_0, \dots, X_n)$ ($n = 0, 1, \dots$). Calcular $Y_n = \mathbf{E}(Z | F_n)$ para $n = 0, 1, \dots$ (a) ¿Existe el límite casi seguro de $\{Y_n\}$?; (b) ¿y el límite en media?

21. Consideremos una sucesión X_1, X_2, \dots , de variables aleatorias independientes dos a dos, y que verifican la siguiente propiedad:

Para cada $n = 1, 2, \dots$, y para cada $j, k = 1, \dots, n$, la distribución del vector aleatorio $(X_1, \dots, X_j, \dots, X_k, \dots, X_n)$ coincide con la del vector aleatorio $(X_1, \dots, X_k, \dots, X_j, \dots, X_n)$.

(a) Demostrar que la sucesión $\{X_n\}$ está formada por variables aleatorias idénticamente distribuidas. (b) Si existe $\mathbf{E} X_1 = a$, demostrar

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow a \quad c.s.,$$

es decir, se verifica la ley fuerte de los grandes números.

⁷Ver Dellacherie, C; Meyer, P.A. Probabilities and potential. North Holland: Amsterdam New York, 1978.

Capítulo 10

Proceso de Poisson y proceso de Wiener

En este capítulo estudiamos *procesos aleatorios con tiempo continuo*, es decir, familias de variables aleatorias $\{X_t\}$ definidas en un espacio de probabilidad común $(\Omega, \mathcal{A}, \mathbf{P})$, cuyo índice t , el *tiempo*, toma valores en el conjunto $[0, \infty)$ o en un intervalo $[0, T]$, donde $T > 0$. Una manera alternativa de ver un proceso aleatorio con tiempo continuo, es considerar fijo cada punto ω del espacio de sucesos elementales Ω , obteniéndose una función $X_t(\omega)$ al variar t . Cada una de estas funciones es una *trayectoria* del proceso aleatorio. Decimos que un proceso aleatorio tiene *trayectorias continuas*, cuando estas funciones son continuas para todo ω , con excepción de un conjunto de probabilidad nula. Decimos que un proceso aleatorio $\{X_t\}$ *parte del origen*, cuando $\mathbf{P}(X_0 = 0) = 1$. A lo largo de este capítulo, por simplicidad, consideramos únicamente procesos que parten del origen.

Decimos que un proceso aleatorio $\{X_t\}$ tiene *incrementos independientes*, cuando para cualquier elección de índices $0 \leq s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n$, las variables aleatorias

$$X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}, \dots, X_{t_n} - X_{s_n}$$

son mutuamente independientes. Decimos que un proceso aleatorio $\{X_t\}$ tiene *incrementos estacionarios* (o también *incrementos homogéneos en el tiempo*), cuando para dos tiempos $0 \leq t < t+h$ arbitrarios, la distribución de la variable aleatoria $X_{t+h} - X_t$ no depende de t . Como el proceso parte del origen, la distribución de las variables aleatorias $X_{t+h} - X_t$ y X_h coinciden.

En el presente capítulo consideramos el proceso de Poisson y el proceso de Wiener, ejemplos básicos de *procesos con incrementos independientes y estacionarios*, también denominados *procesos de Lévy*¹.

10.1. Proceso de Poisson. Definición y caracterizaciones

El *proceso de Poisson* es el ejemplo más sencillo de un proceso con incrementos independientes y estacionarios. Consideremos una sucesión T_1, T_2, \dots de variables aleatorias estrictamente positivas, definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$. Si la variable aleatoria T_k ($k = 1, 2, \dots$) representa la duración de un cierto evento en una serie de eventos consecutivos, entonces, las variables aleatorias

$$S_n = T_1 + \dots + T_n \quad (n = 1, 2, \dots), \quad S_0 = 0 \quad (10.1)$$

representan el tiempo total transcurrido hasta la finalización del n -ésimo evento. El proceso aleatorio $\{N_t\}$ definido mediante

$$N_t = \text{máx}\{n \geq 0: S_n \leq t\} \quad (t \geq 0), \quad (10.2)$$

se denomina *proceso de conteo*, ya que la variable aleatoria N_t cuenta la cantidad de eventos ocurridos hasta el instante t . Observemos que las trayectorias de un proceso de conteo son no decrecientes, constantes en intervalos, toman únicamente valores naturales, y presentan discontinuidades con saltos de amplitud uno en cada uno de los instantes $t = S_n$. Siempre suponemos que $N_t \rightarrow \infty$ ($t \rightarrow \infty$) *c.s.*

Definición 10.1 (Proceso de Poisson). *El proceso de conteo $\{N_t\}$ dado en (10.2) es un proceso de Poisson de parámetro $\alpha > 0$, cuando $\{T_n\}$ es una sucesión de variables aleatorias independientes, idénticamente distribuidas, con distribución exponencial de parámetro α .*

¹Un tratamiento general de los procesos de Lévy puede verse en: J. Bertoin, *Lévy Processes*, Cambridge University Press: Cambridge 1996; K. Sato, *Lévy processes and infinitely divisible distributions*. Cambridge University Press: Cambridge, 1999. A. Kyprianou, *Introductory lectures on fluctuations of Lévy processes with applications*. Springer: Berlin, 2006.

Una variable aleatoria T con distribución exponencial de parámetro $\alpha > 0$ tiene densidad $p(t) = \alpha e^{-\alpha t}$ ($t \geq 0$), $p(t) = 0$ ($t < 0$) (ejemplo 3.8). Veamos otra caracterización de su distribución.

Lema 10.1 (Exponencial). (a) *Consideremos una función $G: [0, \infty) \rightarrow [0, \infty)$ monótona no creciente, no constante, que verifica $G(0) = 1$, tal que existe t con $G(t) > 0$, y que verifica*

$$G(t+h) = G(t)G(h) \text{ para todo } t > 0, h > 0. \quad (10.3)$$

Entonces, existe $\alpha > 0$ tal que $G(t) = e^{-\alpha t}$ ($t \geq 0$).

(b) *Una variable aleatoria $T > 0$ tiene distribución exponencial si y solo si para todo $t > 0, h > 0$, verifica*

$$\mathbf{P}(T > t+h | T > t) = \mathbf{P}(T > h). \quad (10.4)$$

La propiedad (10.4) se denomina pérdida de memoria.

Demostración. (a) Supongamos que $G(1) > 0$ (si se trata de otro punto, la demostración se adapta sin dificultad). Consideremos un número racional $t = p/q > 0$ (p, q naturales). Aplicando q veces la propiedad (10.3), tenemos $G(1) = G(q/q) = G(1/q)^q$. Aplicando p veces la misma propiedad, obtenemos

$$G(t) = G(p/q) = G(1/q)^p = G(1)^{p/q} = G(1)^t.$$

Sabemos que $0 < G(1) \leq 1$. Si $G(1) = 1$, como la función es monótona (no creciente), obtenemos que es constante; luego $\alpha = -\ln G(1) > 0$, y se verifica $G(t) = e^{-\alpha t}$ para todo $t \geq 0$ racional. La propiedad de monotonía permite obtener que $G(t) = e^{-\alpha t}$ para todo real $t \geq 0$, concluyendo la demostración de (a).

Veamos (b). Si una variable aleatoria T tiene distribución exponencial, la fórmula (10.4) es inmediata. Por otra parte, si una variable aleatoria verifica (10.4), la función $G(t) = \mathbf{P}(T > t)$ verifica (10.3) y las demás hipótesis de la parte (a), concluyendo que existe $\alpha > 0$ tal que $G(t) = e^{-\alpha t}$ ($t \geq 0$). Esto equivale a decir que T tiene distribución exponencial. \square

Volvamos al proceso de Poisson. Es claro que del conocimiento de una de las familias de variables aleatorias $\{T_n\}$, $\{S_n\}$, ó $\{N_t\}$, se determinan completamente las otras dos. Comenzamos estudiando la relación entre $\{T_n\}$ y $\{S_n\}$.

Proposición 10.1. *Las variables aleatorias T_1, \dots, T_n son independientes e idénticamente distribuidas, con distribución exponencial de parámetro $\alpha > 0$, si y solo si el vector aleatorio (S_1, \dots, S_n) , definido en (10.1), tiene densidad dada por*

$$p(s_1, \dots, s_n) = \begin{cases} \alpha^n e^{-\alpha s_n}, & \text{si } 0 < s_1 < \dots < s_n, \\ 0, & \text{en otro caso.} \end{cases} \quad (10.5)$$

Demostración. Supongamos primero que T_1, \dots, T_n son variables aleatorias independientes, con distribución común exponencial de parámetro $\alpha > 0$. Esto equivale a decir que el vector aleatorio (T_1, \dots, T_n) tiene densidad $r(t_1, \dots, t_n) = \alpha^n \exp(-\alpha(t_1 + \dots + t_n))$, si $t_k \geq 0$ ($k = 1, \dots, n$), $r(t_1, \dots, t_n) = 0$ en otro caso (proposición 3.3). Sean a_1, \dots, a_n reales positivos arbitrarios. Tenemos

$$\begin{aligned} \mathbf{P}(S_1 \leq a_1, \dots, S_n \leq a_n) &= \mathbf{P}(T_1 \leq a_1, \dots, T_1 + \dots + T_n \leq a_n) \\ &= \int_{\{t_1 \leq a_1, \dots, t_1 + \dots + t_n \leq a_n\}} \alpha^n e^{-\alpha(t_1 + \dots + t_n)} dt_1 \dots dt_n \\ &= \int_0^{a_1} \dots \int_0^{a_n} \mathbf{1}_{\{0 < s_1 < \dots < s_n\}} \alpha^n e^{-\alpha s_n} ds_1 \dots ds_n, \end{aligned}$$

donde en la primer integral múltiple² hicimos el cambio de variables $s_1 = t_1, \dots, s_n = t_1 + \dots + t_n$. Las condiciones que definen los respectivos dominios de integración verifican $\{t_1 \leq a_1, \dots, t_1 + \dots + t_n \leq a_n\} = \{s_1 \leq a_1, \dots, s_n \leq a_n\}$ y $\{t_1 > 0, \dots, t_n > 0\} = \{0 < s_1 < \dots < s_n\}$, el jacobiano del cambio de variable es igual a uno. La identidad obtenida indica que la densidad del vector (S_1, \dots, S_n) es la dada en (10.5), de acuerdo a (3.18).

Veamos el recíproco. Supongamos que el vector aleatorio (S_1, \dots, S_n) tiene densidad dada por (10.5), y sean b_1, \dots, b_n reales arbitrarios. Tenemos

$$\begin{aligned} \mathbf{P}(T_1 \leq b_1, \dots, T_n \leq b_n) &= \mathbf{P}(S_1 \leq b_1, S_2 - S_1 \leq b_2, \dots, S_n - S_{n-1} \leq b_n) \\ &= \int_0^{b_1} ds_1 \int_{s_1}^{s_1 + b_2} ds_2 \dots \int_{s_{n-1}}^{s_{n-1} + b_n} \alpha^n e^{-\alpha s_n} ds_n \\ &= (1 - e^{-\alpha b_1}) \dots (1 - e^{-\alpha b_n}), \end{aligned}$$

²Abreviamos $\int_0^\infty \dots \int_0^\infty \mathbf{1}_B f(t_1, \dots, t_n) dt_1 \dots dt_n$ por $\int_B f(t_1, \dots, t_n) dt_1 \dots dt_n$.

donde, para obtener la última igualdad, calculamos la integral iterada. De aquí, tomando $b_j \rightarrow \infty$ ($j \neq k, j = 1, \dots, n$) resulta que $\mathbf{P}(T_k \leq b_k) = 1 - e^{-\alpha b_k}$ ($k = 1, \dots, n$), y las variables aleatorias tienen distribución exponencial de parámetro α . Entonces, se verifica también la propiedad (3.24) que define su independencia. Esto concluye la demostración. \square

Antes de seguir es útil calcular la siguiente integral múltiple. Considerando dos reales $0 \leq a < b$, tenemos

$$\int_{\{a < s_1 < \dots < s_n < b\}} ds_1 \cdots ds_n = \int_a^b ds_n \int_a^{s_n} ds_{n-1} \cdots \int_a^{s_2} ds_1 = \frac{(b-a)^n}{n!}, \quad (10.6)$$

donde calculamos, una a una, las n integrales simples que componen la integral iterada. El mismo resultado se obtiene también, observando que la integral múltiple en (10.6) es la $n!$ -ésima parte del hipercubo $[a, b]^n$ en \mathbb{R}^n .

La proposición 10.1 nos permite obtener la distribución de las variables aleatorias S_n ($n = 1, 2, \dots$) y N_t ($t \geq 0$), en un proceso de Poisson.

Corolario 10.1. *Consideremos un proceso de Poisson $\{N_t\}$ de parámetro $\alpha > 0$.*

(a) *Para cada $n = 1, 2, \dots$ la variable aleatoria $S_n = T_1 + \dots + T_n$ tiene densidad dada por*

$$q(x) = \alpha^n \frac{x^{n-1}}{(n-1)!} e^{-\alpha x}, \quad (x > 0), \quad (10.7)$$

$q(x) = 0$ ($x \leq 0$), y función de distribución dada por

$$F(x) = 1 - e^{-\alpha x} \left(1 + \alpha x + \frac{(\alpha x)^2}{2!} + \dots + \frac{(\alpha x)^{n-1}}{(n-1)!} \right), \quad (10.8)$$

denominada distribución de Erlang de parámetros (α, n) .

(b) *Para cada $t > 0$ la variable aleatoria N_t tiene distribución de Poisson de parámetro αt , es decir*

$$\mathbf{P}(N_t = n) = e^{-\alpha t} (\alpha t)^n / n! \quad (n = 0, 1, \dots).$$

Observación. La distribución de Erlang con parámetros (α, n) es un caso particular de la distribución Gama con parámetros (α, λ) (introducida en el ejemplo 3.12), cuando el que el parámetro λ toma valores naturales y positivos.

Demostración. (a) Utilizando la fórmula (10.5), tenemos

$$\begin{aligned} \mathbf{P}(S_n \leq t) &= \int_0^t e^{-\alpha s_n} \left[\int_0^{s_n} ds_{n-1} \cdots \int_0^{s_2} ds_1 \right] ds_n \\ &= \int_0^t e^{-\alpha s_n} \frac{s_n^{n-1}}{(n-1)!} ds_n, \end{aligned}$$

donde en la integral múltiple $(n-1)$ -dimensional entre paréntesis rectos utilizamos el cálculo en (10.6). Esto demuestra (10.7).

Una demostración alternativa de (a) se basa en el teorema 6.2 de unicidad de la función característica. Por un lado tenemos

$$\mathbf{E} e^{i\mu S_n} = \mathbf{E} e^{i\mu(T_1 + \cdots + T_n)} = \prod_{k=1}^n \mathbf{E} e^{i\mu T_k} = \left(\frac{\alpha}{\alpha - i\mu} \right)^n, \quad (10.9)$$

donde utilizamos la independencia y la fórmula de la función característica de una variable aleatoria con distribución exponencial, calculada en el ejemplo 6.5. Por otra parte, mediante el cambio de variable $y = (\alpha - i\mu)x$, obtenemos

$$\begin{aligned} \mathbf{E} e^{i\mu S_n} &= \int_0^\infty e^{i\mu x} \alpha^n e^{-\alpha x} \frac{x^{n-1}}{(n-1)!} dx = \alpha^n \int_0^\infty e^{-(\alpha - i\mu)x} \frac{x^{n-1}}{(n-1)!} dx \\ &= \left(\frac{\alpha}{\alpha - i\mu} \right)^n \int_0^\infty e^{-y} \frac{y^{n-1}}{(n-1)!} dy = \left(\frac{\alpha}{\alpha - i\mu} \right)^n, \end{aligned} \quad (10.10)$$

donde la última integral se calcula a partir de la función Gama (ver ejemplo 3.12). En conclusión, la igualdad de resultados en (10.9) y (10.10) nos permite concluir la fórmula de la densidad (10.7). Para verificar la fórmula (10.8), la derivamos con respecto de x , obteniendo la fórmula de la densidad (10.7).

Veamos (b). Tenemos la igualdad de sucesos $\{N_t = n\} = \{S_n \leq t < S_{n+1}\}$. Como las variables aleatorias S_n y T_{n+1} son independientes, utilizando la fórmula de la densidad (10.7), tenemos

$$\begin{aligned} \mathbf{P}(N_t = n) &= \mathbf{P}(S_n \leq t < S_n + T_{n+1}) \\ &= \int_0^t \alpha^n e^{-\alpha s} \frac{s^{n-1}}{(n-1)!} ds \int_{t-s}^\infty \alpha e^{-\alpha u} du \\ &= \int_0^t \alpha^n \frac{s^{n-1}}{(n-1)!} e^{-\alpha t} ds = e^{-\alpha t} (\alpha t)^n / n!, \end{aligned}$$

concluyendo la demostración. \square

Consideremos un proceso de Poisson $\{N_t\}$. Queremos ahora calcular, dados índices arbitrarios $0 < t_1 < \dots < t_k$, la distribución del vector aleatorio $(N_{t_1}, \dots, N_{t_k})$, que denominamos *distribución finito-dimensional* de proceso aleatorio. Como la distribución de este vector es discreta (y el proceso tiene trayectorias no decrecientes), es claro que es suficiente determinar, dados naturales arbitrarios $0 = m_0 \leq m_1 \leq \dots \leq m_k$, la probabilidad

$$\begin{aligned} \mathbf{P}(N_{t_1} = m_1, N_{t_2} = m_2, \dots, N_{t_k} = m_k) \\ = \mathbf{P}(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2, \dots, N_{t_k} - N_{t_{k-1}} = n_k), \end{aligned}$$

donde designamos $n_j = m_j - m_{j-1}$ ($j = 1, \dots, k$). Para calcular esta última probabilidad demostramos que los incrementos de un proceso de Poisson son independientes y estacionarios, y calculamos su distribución.

Teorema 10.1. *Un proceso de conteo $\{N_t\}$ es un proceso de Poisson de parámetro $\alpha > 0$, si y solo si se verifican las siguientes propiedades:*

(a) *Dados $k \geq 1$, tiempos $0 = t_0 \leq t_1 < \dots < t_k$ y naturales n_1, \dots, n_k , todos arbitrarios, se verifica*

$$\mathbf{P}(N_{t_1} = n_1, \dots, N_{t_k} - N_{t_{k-1}} = n_k) = \prod_{j=1}^k \mathbf{P}(N_{t_j} - N_{t_{j-1}} = n_j).$$

(b) *Dados tiempos $0 \leq t < t + h$ arbitrarios, se verifica*

$$\mathbf{P}(N_{t+h} - N_t = n) = e^{-\alpha h} (\alpha h)^n / n! \quad (n = 0, 1, \dots).$$

Observación. La propiedad (a) es equivalente a la independencia de incrementos, como resulta de la aplicación de la proposición 3.2.

Demostración. Designemos $m_j = n_1 + \dots + n_j$ ($j = 1, \dots, k$). De las proposiciones (a) y (b), obtenemos la propiedad

$$\mathbf{P}\left(\bigcap_{j=1}^k \{N_{t_j} - N_{t_{j-1}} = n_j\}\right) = \prod_{j=1}^k e^{-\alpha(t_j - t_{j-1})} \frac{(\alpha(t_j - t_{j-1}))^{n_j}}{n_j!}. \quad (10.11)$$

Veamos también que esta propiedad implica (a) y (b). Tomando en (10.11) $k = 2$, y los valores $t_1 = t$, $t_2 = t + h$, $n_2 = n$, y sumando para $n_1 = 0, 1, \dots$, obtenemos la fórmula en (b); que sustituida en (10.11) nos permite obtener (a). Hemos entonces demostrado que (a) y (b) son equivalentes a (10.11).

Supongamos ahora que $\{N_t\}$ es un proceso de Poisson de parámetro $\alpha > 0$, y verifiquemos (10.11). Tenemos que calcular la probabilidad del suceso

$$\mathbf{A} = \bigcap_{j=1}^k \{N_{t_j} - N_{t_{j-1}} = n_j\} = \bigcap_{j=1}^k \{N_{t_j} = m_j\} = \bigcap_{j=1}^k \{S_{m_j} \leq t_j < S_{m_j+1}\}.$$

El cálculo es similar al de la demostración de (b) en el corolario 10.1. Si designamos $B_j = \{t_{j-1} < s_{m_{j-1}+1} < \dots < s_{m_j} \leq t_j\}$ ($j = 1, \dots, k$), tenemos $\mathbf{A} = \{(S_1, \dots, S_{m_k}) \in \bigcap_{j=1}^k B_j\} \cap \{t_k < S_{m_k} + T_{m_k+1}\}$ y podemos, como las variables aleatorias S_{m_k} y T_{m_k+1} son independientes, calcular la probabilidad

$$\begin{aligned} \mathbf{P}(\mathbf{A}) &= \int_{\bigcap_{j=1}^k B_j} q(s_1, \dots, s_{m_k}) ds_1 \cdots ds_{m_k} \int_{t_k - s_{m_k}}^{\infty} \alpha e^{-\alpha u} du \\ &= \alpha^{m_k} e^{-\alpha t_k} \int_{\bigcap_{j=1}^k B_j} ds_1 \cdots ds_{m_k} = \prod_{j=1}^k e^{-\alpha(t_j - t_{j-1})} \frac{(\alpha(t_j - t_{j-1}))^{n_j}}{n_j!}, \end{aligned}$$

donde $q(s_1, \dots, s_{m_k})$ es la densidad del vector (S_1, \dots, S_{m_k}) , y la última integral es el producto de k integrales (en los conjuntos B_j , $j = 1, \dots, k$) cada una de las cuales se calcula mediante la fórmula (10.6). Esto concluye la demostración de la primera parte.

Veamos la demostración del recíproco, suponiendo que un proceso de conteo $\{N_t\}$ verifica (10.11). Consideremos los puntos $0 \leq r_1 < t_1 < \dots < r_n < t_n$, y los intervalos $I_j = (r_j, t_j]$ ($j = 1, \dots, n$). Tenemos

$$\begin{aligned} &\mathbf{P}(r_1 < S_1 \leq t_1, \dots, r_n < S_n \leq t_n) \\ &= \mathbf{P}(N_{r_1} = 0, N_{t_1} - N_{r_1} = 1, \dots, N_{r_n} - N_{t_{n-1}} = 0, N_{t_n} - N_{r_n} \geq 1) \\ &= \alpha^{n-1} (t_1 - r_1) \cdots (t_{n-1} - r_{n-1}) (1 - e^{-\alpha(t_n - r_n)}) e^{-\alpha r_n} \\ &= \int_{I_1 \times \dots \times I_n} p(s_1, \dots, s_n) ds_1 \cdots ds_n. \quad (10.12) \end{aligned}$$

donde $p(s_1, \dots, s_n)$ es la densidad dada en (10.5), y la última igualdad se obtiene calculando la integral múltiple. Esta igualdad implica la validez de (3.18) para la densidad del vector (S_1, \dots, S_n) . En vista de la proposición 10.1, como n es arbitrario, concluimos la demostración. \square

Veremos a continuación otra caracterización del proceso de Poisson, que además de su relevancia matemática, explica porque este proceso es adecuado para modelar la ocurrencia de eventos consecutivos en diversas aplicaciones. Comencemos observando que un proceso de Poisson $\{N_t\}$ verifica

$$\begin{aligned}\mathbf{P}(N_t \geq 2) &= 1 - \mathbf{P}(N_t = 0) - \mathbf{P}(N_t = 1) \\ &= 1 - e^{-\alpha t} - \alpha t e^{-\alpha t} = o(t) \quad (t \rightarrow 0).\end{aligned}$$

Esta sencilla propiedad en un proceso de conteo con incrementos independientes y estacionarios, lo caracteriza como un proceso de Poisson.

Teorema 10.2. *Consideremos un proceso de conteo $\{N_t\}$. El proceso aleatorio $\{N_t\}$ es un proceso de Poisson, si y solo si tiene incrementos independientes y estacionarios, y verifica la propiedad:*

$$\mathbf{P}(N_t \geq 2) = o(t) \quad (t \rightarrow 0). \quad (10.13)$$

Demostración. Hemos visto que el proceso de Poisson verifica la propiedad (10.13). Supongamos entonces que un proceso de conteo $\{N_t\}$ tiene incrementos independientes y estacionarios, y verifica (10.13). Según el teorema 10.1, dado que los incrementos son estacionarios, para obtener que $\{N_t\}$ es un proceso de Poisson, es suficiente demostrar que existe $\alpha > 0$ tal que se verifica $\mathbf{P}(N_t = n) = e^{-\alpha t}(\alpha t)^n/n!$ ($n = 0, 1, \dots$). Veamos primero que existe $\alpha > 0$ tal que $\mathbf{P}(N_t = 0) = e^{-\alpha t}$. En efecto, tenemos

$$\mathbf{P}(N_{t+s} = 0) = \mathbf{P}(N_{t+s} - N_t = 0, N_t = 0) = \mathbf{P}(N_s = 0) \mathbf{P}(N_t = 0),$$

dado que los incrementos son independientes y estacionarios. Como la función $G(t) = \mathbf{P}(N_t = 0) = \mathbf{P}(T_1 > t)$ es no creciente, no constante (porque $G(t) \rightarrow 0$ si $t \rightarrow \infty$), y $\mathbf{P}(T_1 > t)$ es positivo para algún $t > 0$ (si $\mathbf{P}(T_1 > t) = 0$ para todo $t > 0$, entonces $\mathbf{P}(T_1 = 0) = 1$), aplicando el lema 10.1 obtenemos, que existe un real $\alpha > 0$ tal que $\mathbf{P}(N_t = 0) = e^{-\alpha t}$. Sabemos además, que

$$\begin{aligned}\mathbf{P}(N_t = 1) &= 1 - \mathbf{P}(N_t = 0) - \mathbf{P}(N_t \geq 2) \\ &= 1 - e^{-\alpha t} + o(t) = \alpha t + o(t) \quad (t \rightarrow 0).\end{aligned}$$

Consideremos ahora para cada $z \in (0, 1)$ la función $H(t)$ ($t \geq 0$), definida mediante

$$H(t) = \mathbf{E} z^{N_t} = \mathbf{P}(N_t = 0) + z \mathbf{P}(N_t = 1) + z^2 \mathbf{P}(N_t = 2) + \dots$$

En vista de la independencia y la estacionariedad de los incrementos de $\{N_t\}$, obtenemos

$$H(t+s) = \mathbf{E} z^{N_{t+s}} = \mathbf{E} z^{N_{t+s}-N_s} \mathbf{E} z^{N_s} = H(t)H(s).$$

Es sencillo de ver, dado que los procesos de conteo verifican $N_t \rightarrow \infty$ ($t \rightarrow \infty$) *c.s.*, que la función $H(t)$ verifica las hipótesis del lema 10.1, de donde obtenemos que para cada z existe una constante $h(z)$, tal que se verifica

$$H(t) = e^{h(z)t}. \quad (10.14)$$

Determinemos ahora $h(z)$ calculando la derivada de la función $H(t)$ en el punto $t = 0$. Tenemos

$$H(t) - H(0) = e^{-\alpha t} - 1 + z \mathbf{P}(N_t = 1) + \mathbf{E} z^{N_t} \mathbf{1}_{\{N_t \geq 2\}}.$$

Como $|\mathbf{E} z^{N_t} \mathbf{1}_{\{N_t \geq 2\}}| \leq \mathbf{P}(N_t \geq 2) = o(t)$, y $\mathbf{P}(N_t = 1) = \alpha t + o(t)$, se verifica

$$h(z) = H'(0) = \lim_{t \rightarrow 0} \frac{H(t) - H(0)}{t} = -\alpha + \alpha z.$$

Obtenemos finalmente, desarrollando la función $e^{\alpha t z}$ en serie de potencias, la identidad

$$\mathbf{E} z^{N_t} = \sum_{n=0}^{\infty} z^n \mathbf{P}(N_t = n) = e^{-\alpha t + \alpha t z} = \sum_{n=0}^{\infty} e^{-\alpha t} \frac{(\alpha t)^n}{n!} z^n.$$

Identificando los coeficientes de igual grado en las dos series de potencias, obtenemos que N_t tiene distribución de Poisson con parámetro αt , concluyendo la demostración. \square

10.2. Proceso de Poisson compuesto y aplicaciones

Consideremos una sucesión $T_1, Z_1, T_2, Z_2, \dots$ de variables aleatorias independientes. Supongamos que las variables aleatorias $\{T_n\}$ son idénticamente distribuidas, con distribución exponencial de parámetro $\alpha > 0$, y sea $\{N_t\}$ el proceso de Poisson definido en (10.2). Por su parte, las variables

aleatorias $\{Z_n\}$ son también idénticamente distribuidas. Consideremos, para cada $t \geq 0$, la variable aleatoria

$$Y_t = \sum_{k=1}^{N_t} Z_k,$$

donde entendemos $Y_t = 0$ si $N_t = 0$, que es la suma de una cantidad aleatoria de sumandos. El proceso aleatorio $\{Y_t\}$ se llama *proceso de Poisson compuesto*. Sus trayectorias son constantes en los intervalos en los que $\{N_t\}$ es constante, y si S_n denota el n -ésimo salto de proceso de Poisson, la magnitud del salto de $\{Y_t\}$ en el instante $t = S_n$ es Z_n .

En matemática actuarial se considera el siguiente modelo para la evolución del capital de una compañía de seguros. Definimos, para cada $t \geq 0$, la variable aleatoria

$$X_t = x + ct - \sum_{k=1}^{N_t} Z_k. \quad (10.15)$$

El proceso aleatorio $\{X_t\}$, que modela el capital de la compañía, se denomina *proceso de riesgo*. El capital inicial $X_0 = x$ es un real positivo, la constante $c > 0$ es la *tasa de pago* de los seguros, es decir, suponemos que la compañía recibe un monto ch en cada intervalo de tiempo $[t, t+h]$. En cada uno de los instantes S_n ($n = 1, 2, \dots$) esta compañía debe pagar un reclamo de un monto Z_n (que suponemos positivo). Es importante entonces, conocer la magnitud

$$\mathbf{P}(\exists t \geq 0: X_t \leq 0),$$

la probabilidad de que la compañía tenga un capital “negativo”, que llamamos *probabilidad de ruina*.

En general no es posible calcular explícitamente esta magnitud. Sin embargo, en el caso particular en el que los reclamos $\{Z_n\}$ tienen distribución exponencial, la probabilidad de ruina se calcula en forma exacta, como vemos a continuación.

Teorema 10.3. *Consideremos el proceso de riesgo $\{X_t\}$ en (10.15), con reclamos $\{Z_n\}$ con distribución exponencial de parámetro $\beta > 0$, que verifica $\alpha < c\beta$. Entonces, la probabilidad de ruina está dada por*

$$\mathbf{P}(\exists t \geq 0: X_t \leq 0) = \frac{\alpha}{c\beta} e^{-(\beta - \alpha/c)x}, \quad x > 0.$$

Demostración. Como las trayectorias del proceso $\{X_t\}$ son crecientes en los intervalos entre los instantes de salto, el primer t tal que $X_t \leq 0$ es un instante $t = S_n$, es decir, un instante un salto de la trayectoria. Tenemos entonces

$$\mathbf{P}(\exists t \geq 0: X_t \leq 0) = \mathbf{P}(\exists n \geq 1: X_{S_n} \leq 0). \quad (10.16)$$

El valor del proceso en estos instantes, es

$$\begin{aligned} X_{S_n} &= x + cS_n - \sum_{k=1}^n Z_k \\ &= x + \sum_{k=1}^n (cT_k - Z_k) = x + U_n, \end{aligned} \quad (10.17)$$

donde introducimos la notación $U_n = \sum_{k=1}^n V_k$ ($n = 1, 2, \dots$), y designamos $V_k = cT_k - Z_k$ ($k = 1, 2, \dots$). La sucesión $\{U_n\}$ se denomina *paseo al azar asociado* al proceso de riesgo $\{X_t\}$, y en vista de (10.16) y (10.17), tenemos

$$\mathbf{P}(\exists t \geq 0: X_t \leq 0) = \mathbf{P}(\exists n \geq 1: x + U_n \leq 0).$$

Esto significa que calcular la probabilidad de ruina es equivalente a resolver un problema de barrera para el paseo al azar asociado. Aplicando la ley fuerte de los grandes números (teorema 9.7), obtenemos que $U_n/n \rightarrow \mathbf{E}V_1 = c/\alpha - 1/\beta > 0$ ($n \rightarrow \infty$) *c.s.* Por eso $U_n \rightarrow \infty$ ($n \rightarrow \infty$) *c.s.*, y el problema de calcular la probabilidad de ruina consiste en cuantificar la proporción de trayectorias del paseo al azar que alcanzan el nivel $y = 0$, antes de tomar valores grandes.

No es difícil ver que las variables aleatorias cT_k ($k = 1, 2, \dots$) tienen distribución exponencial de parámetro $\gamma = \alpha/c$. Además, en el ejemplo 6.8 vimos, que las variables aleatorias V_k ($k = 1, 2, \dots$) tienen densidad dada por

$$p(y) = \begin{cases} \frac{\beta\gamma}{\beta+\gamma} e^{-\gamma y}, & \text{si } y > 0, \\ \frac{\beta\gamma}{\beta+\gamma} e^{\beta y}, & \text{si } y \leq 0. \end{cases}$$

Consideremos ahora la función auxiliar

$$R(x) = \begin{cases} \frac{\alpha}{c\beta} e^{-(\beta-\alpha/c)x}, & \text{si } x \geq 0, \\ 1, & \text{si } x < 0. \end{cases} \quad (10.18)$$

Veamos que para $x > 0$, se verifica la siguiente ecuación³:

$$\begin{aligned} \int_{-\infty}^{\infty} R(x+y)p(y)dy &= \frac{\beta\gamma}{\beta+\gamma} \int_{-\infty}^{-x} e^{\beta y} dy + \frac{\gamma^2}{\beta+\gamma} \int_{-x}^0 e^{-(\beta-\gamma)(x+y)} e^{\beta y} dy \\ &+ \frac{\gamma^2}{\beta+\gamma} \int_0^{\infty} e^{-(\beta-\gamma)(x+y)} e^{-\gamma y} dy = \frac{\gamma}{\beta} e^{-(\beta-\gamma)x} = R(x). \end{aligned} \quad (10.19)$$

(Esta igualdad es equivalente a $\mathbf{E} R(x + V_1) = R(x)$, con $x > 0$.)

Consideremos los vectores aleatorios $F_n = (V_1, \dots, V_n)$ ($n = 1, 2, \dots$), y el tiempo de parada respecto de $\{F_n\}$ definido mediante

$$\tau = \inf\{n \geq 0: x + U_n \leq 0\},$$

y veamos que la sucesión $\{R(x + U_{\tau \wedge n})\}$ es una martingala, adaptada a $\{F_n\}$. En efecto,

$$\begin{aligned} \mathbf{E}(R(x + U_{\tau \wedge (n+1)}) | F_n) &= \mathbf{E}(R(x + U_{\tau}) \mathbf{1}_{\{\tau \leq n\}} | F_n) + \mathbf{E}(R(x + U_n + V_{n+1}) \mathbf{1}_{\{\tau > n\}} | F_n) \\ &= R(x + U_{\tau}) \mathbf{1}_{\{\tau \leq n\}} + \mathbf{1}_{\{\tau > n\}} \int_{-\infty}^{\infty} R(x + U_n + y)p(y)dy \\ &= R(x + U_{\tau}) \mathbf{1}_{\{\tau \leq n\}} + \mathbf{1}_{\{\tau > n\}} R(x + U_n) = R(x + U_{\tau \wedge n}). \end{aligned}$$

Aquí nos hemos basado en los siguientes hechos: (a) las variables aleatorias $R(x + U_{\tau}) \mathbf{1}_{\{\tau \leq n\}}$ y $\mathbf{1}_{\{\tau > n\}}$ son funciones del vector aleatorio F_n ; (b) calculamos la tercer esperanza condicional como en el ejemplo 9.6; (c) en el suceso $\{\tau > n\}$, vale $x + U_n > 0$, y podemos aplicar aplicar la identidad (10.19) para obtener la última igualdad. Estamos en condiciones de aplicar el teorema 9.3 del muestreo opcional, con el tiempo de parada $\tau \wedge n$. Observando que $R(x + U_{\tau}) \mathbf{1}_{\{\tau \leq n\}} = \mathbf{1}_{\{\tau \leq n\}}$, tenemos

$$R(x) = \mathbf{E} R(x + U_{\tau \wedge n}) = \mathbf{P}(\tau \leq n) + \mathbf{E} R(x + U_n) \mathbf{1}_{\{\tau > n\}}. \quad (10.20)$$

Como $U_n \rightarrow \infty$ ($n \rightarrow \infty$) *c.s.*, y $R(x) \rightarrow 0$ ($x \rightarrow \infty$), se verifica $R(x + U_n) \rightarrow 0$ ($n \rightarrow \infty$), *c.s.* Obtenemos entonces $\mathbf{E} R(x + U_n) \mathbf{1}_{\{\tau > n\}} \rightarrow 0$ ($n \rightarrow \infty$), porque la función $R(x)$ es acotada. Al tomar límite si $n \rightarrow \infty$ en la fórmula (10.20), para $x > 0$, obtenemos

$$R(x) = \mathbf{P}(\tau < \infty) = \mathbf{P}(\exists n \geq 1: x + U_n \leq 0).$$

En vista de la definición de $R(x)$ en (10.18), esto concluye la demostración.

□

³La ecuación integral (10.19) se llama *ecuación de Wiener-Hopf*, ver ejercicio 12 del capítulo 9.

10.3. Proceso de Wiener. Definición y primeras propiedades

En esta sección estudiamos el *proceso de Wiener*, también denominado *movimiento Browniano*. Este proceso aleatorio es el ejemplo básico de diversas familias de procesos aleatorios (entre ellas: procesos de Markov y martingalas con tiempo continuo, procesos de Lévy), y es frecuentemente utilizado en la modelización matemática en las más diversas ciencias (por ejemplo: física, biología, finanzas), jugando también un importante rol en la estadística matemática.

Su primera denominación se debe a las investigaciones del botánico inglés Robert Brown, que en 1828 observó y describió el movimiento caótico de una partícula de polen suspendida en agua, destacando la naturaleza física (y no biológica) del movimiento observado. Luego de las contribuciones de L. Bachelier (1900), quién propuso este modelo para las fluctuaciones de la bolsa de París, y de A. Einstein (1905) y M. Smoluchowski (1906), que lo propusieron en el marco de la teoría molecular de la materia; Norbert Wiener, en 1923, construyó el proceso aleatorio con trayectorias continuas correspondiente a la dinámica observada, y a la definición que presentamos a continuación.

Definición 10.2. *Un proceso aleatorio $\{W_t\}$ es un proceso de Wiener, si se verifican las siguientes propiedades:*

- (a) *El proceso parte del origen, es decir, $\mathbf{P}(W_0 = 0) = 1$.*
- (b) *Las trayectorias de $\{W_t\}$ son funciones continuas.*
- (c) *El proceso aleatorio $\{W_t\}$ tiene incrementos independientes.*
- (d) *Dados $0 \leq t < t + h$, la variable aleatoria $W_{t+h} - W_t$ tiene distribución normal, con esperanza nula, y varianza $\mathbf{var}(W_{t+h} - W_t) = h$.*

Una de las características más interesantes del proceso de Wiener es la naturaleza de sus trayectorias. Consideramos un intervalo $[0, T]$, y una sucesión de particiones

$$\lambda^n = \{0 = t_0^n < t_1^n < \cdots < t_{k(n)}^n = T\} \quad (n = 1, 2, \dots), \quad (10.21)$$

cuya norma $|\lambda^n| = \max\{t_k^n - t_{k-1}^n : k = 1, \dots, k(n)\}$ tiende a cero si $n \rightarrow \infty$, y tales que se verifica $\lambda^n \subset \lambda^{n+1}$ ($n = 1, 2, \dots$) es decir, cada partición se

obtiene de la anterior agregando puntos. Para una función $f: [0, T] \rightarrow \mathbb{R}$ con derivada $f'(t)$ continua en $[0, T]$, cuando $n \rightarrow \infty$, tenemos

$$\sum_{k=1}^{k(n)} |f(t_k^n) - f(t_{k-1}^n)| = \sum_{k=1}^{k(n)} |f'(\theta_k^n)|(t_k^n - t_{k-1}^n) \rightarrow \int_0^T |f'(t)| dt.$$

(Aquí aplicamos el teorema del valor medio, $\theta_k^n \in [t_{k-1}^n, t_k^n]$ para cada n y cada k .) El límite obtenido es la *variación* de la función $f(t)$ en el intervalo $[0, T]$. En forma similar, si $|f'(t)| \leq M$ ($0 \leq t \leq T$), tenemos

$$\sum_{k=1}^{k(n)} (f(t_k^n) - f(t_{k-1}^n))^2 \leq \sum_{k=1}^{k(n)} M^2 (t_k^n - t_{k-1}^n)^2 \leq M^2 T |\lambda^n| \rightarrow 0,$$

si $n \rightarrow \infty$, y decimos, que la *variación cuadrática* de la función $f(t)$ en el intervalo $[0, T]$ es nula.

El siguiente teorema muestra que las trayectorias de un proceso de Wiener presentan un comportamiento diferente: su variación en un intervalo $[0, T]$ no existe, es infinita; y su variación cuadrática en un intervalo $[0, T]$ es igual a T .

Teorema 10.4 (Propiedades de las trayectorias).

Consideremos un proceso de Wiener $\{W_t\}$ y una sucesión creciente de particiones $\{\lambda^n\}$ como en (10.21), cuyas normas $\{|\lambda^n|\}$ tienden a cero si $n \rightarrow \infty$. Se verifica

$$V_n = \sum_{k=1}^{k(n)} |W_{t_k^n} - W_{t_{k-1}^n}| \rightarrow \infty \quad (n \rightarrow \infty) \quad c.s. \quad (10.22)$$

$$Q_n = \sum_{k=1}^{k(n)} (W_{t_k^n} - W_{t_{k-1}^n})^2 \rightarrow T \quad (n \rightarrow \infty) \quad \text{en media cuadrática.} \quad (10.23)$$

Además, si $\sum_{n=1}^{\infty} |\lambda^n| < \infty$, la convergencia en (10.23) es casi segura.

Demostración. Comencemos con la demostración de (10.22). En primer lugar observemos que como las particiones son crecientes, aplicando la propiedad triangular, se obtiene que $V_n \leq V_{n+1}$ ($n = 1, 2, \dots$), es decir, la sucesión $\{V_n\}$ es no decreciente, casi seguramente. Queremos demostrar

que $\mathbf{P}(V_n \rightarrow \infty) = 1$. Esto es equivalente a demostrar que dado $K > 0$ arbitrario, se verifica $\mathbf{P}(\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{V_m(\omega) > K\}) = 1$. Esta igualdad, tomando complementos, es equivalente a $\mathbf{P}(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} \{V_m(\omega) \leq K\}) = 0$. Como la sucesión $\{V_n\}$ es no decreciente, tenemos

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{V_m(\omega) \leq K\}\right) \leq \mathbf{P}\left(\bigcup_{m=n}^{\infty} \{V_m(\omega) \leq K\}\right) = \mathbf{P}(V_n(\omega) \leq K).$$

En conclusión, para demostrar (10.22), verificamos que $\mathbf{P}(V_n(\omega) \leq K) \rightarrow 0$ ($n \rightarrow \infty$). Como $\{W_t\}$ tiene incrementos independientes, aplicando la fórmula (4.19), obtenemos

$$\mathbf{var} V_n = \sum_{k=1}^{k(n)} \mathbf{var} |W_{t_k^n} - W_{t_{k-1}^n}| \leq \sum_{k=1}^{k(n)} \mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}|^2 = T.$$

Por otra parte, tenemos $\mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}| = \sqrt{t_k^n - t_{k-1}^n} \mathbf{E} |Z|$, donde Z es una variable aleatoria con distribución normal estándar, y $\mathbf{E} |Z| = \sqrt{2/\pi}$. Entonces, como $\sqrt{t_k^n - t_{k-1}^n} \geq (t_k^n - t_{k-1}^n)/\sqrt{|\lambda^n|}$, tenemos

$$\mathbf{E} V_n = \sum_{k=1}^{k(n)} \mathbf{E} |W_{t_k^n} - W_{t_{k-1}^n}| = \mathbf{E} |Z| \sum_{k=1}^{k(n)} \sqrt{t_k^n - t_{k-1}^n} \geq \frac{T \mathbf{E} |Z|}{\sqrt{|\lambda^n|}} \rightarrow \infty,$$

si $n \rightarrow \infty$. Para n suficientemente grande se verifica $\mathbf{E} V_n > K$, y aplicando la desigualdad de Chebishev (4.21), obtenemos

$$\begin{aligned} \mathbf{P}(V_n \leq K) &\leq \mathbf{P}(|V_n - \mathbf{E} V_n| \geq \mathbf{E} V_n - K) \\ &\leq \frac{1}{(\mathbf{E} V_n - K)^2} \mathbf{var} V_n \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned}$$

concluyendo la demostración de (10.22).

Veamos ahora la demostración de (10.23). Las variables aleatorias $Y_k = (W_{t_k^n} - W_{t_{k-1}^n})^2 - (t_k^n - t_{k-1}^n)$ ($k = 1, \dots, k(n)$) son independientes, y verifican

$$\mathbf{E} Y_n = 0, \quad \mathbf{var} Y_n = (t_k^n - t_{k-1}^n)^2 \mathbf{E}(Z^2 - 1)^2,$$

si Z designa una variable aleatoria con distribución normal estándar. Aplicando la fórmula (4.19) (los momentos de orden dos coinciden con las

varianzas), tenemos

$$\begin{aligned} \mathbf{E}(Q_n - T)^2 &= \mathbf{E}\left(\sum_{k=1}^{k(n)} Y_n\right)^2 = \sum_{k=1}^{k(n)} \mathbf{E}(Y_n)^2 = \sum_{k=1}^{k(n)} (t_k^n - t_{k-1}^n)^2 \mathbf{E}(Z^2 - 1)^2 \\ &\leq T|\lambda^n| \mathbf{E}(Z^2 - 1)^2 \rightarrow 0 \quad (n \rightarrow \infty), \end{aligned} \quad (10.24)$$

obteniendo la convergencia en media cuadrática.

La convergencia casi segura, bajo el supuesto $\sum_{n=1}^{\infty} |\lambda^n| < \infty$, se obtiene de la siguiente forma. Sabemos que $\mathbf{P}(Q_n \rightarrow T) = 1$ si (y solo si) para todo $\varepsilon > 0$, se cumple $\mathbf{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\}\right) = 0$ (ver fórmula (5.1)). Tenemos

$$\begin{aligned} \mathbf{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\}\right) &\leq \mathbf{P}\left(\bigcup_{n=m}^{\infty} \{|Q_n - T| > \varepsilon\}\right) \\ &\leq \sum_{n=m}^{\infty} \mathbf{P}(|Q_n - T| > \varepsilon) \\ &\leq \frac{1}{\varepsilon^2} \sum_{n=m}^{\infty} \mathbf{E}|Q_n - T|^2 \leq t \mathbf{E}(Z^2 - 1) \sum_{n=m}^{\infty} |\lambda^n| \rightarrow 0, \end{aligned}$$

si $m \rightarrow \infty$, donde utilizamos la acotación obtenida en (10.24). De aquí se obtiene la convergencia casi segura⁴, concluyendo la demostración. \square

10.4. Problemas de barrera para el proceso de Wiener

En esta sección consideramos un *proceso de Wiener con tendencia*, que designamos $\{X_t\}$, y definimos mediante

$$X_t = W_t + at \quad (t \geq 0). \quad (10.25)$$

En (10.25) tenemos un proceso de Wiener $\{W_t\}$, definido en un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbf{P})$, y un real arbitrario a .

Es sencillo de ver que el proceso aleatorio $\{X_t\}$ tiene incrementos independientes y estacionarios, y cumple la siguiente *propiedad de Markov*:

⁴Hemos demostrado que $\sum_{n=1}^{\infty} \mathbf{P}(\mathbf{A}_n) < \infty$ implica $\mathbf{P}(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \mathbf{A}_n) = 0$, que es la primer parte del llamado *lema de Borel–Cantelli*.

Dados $0 < t_1 < \dots < t_n = t < t + h$ y una función $h(x)$ tal que existe $\mathbf{E}h(X_{t+h})$, se verifica

$$E(h(X_{t+h}) | X_t, \dots, X_{t_1}) = E(h(X_{t+h}) | X_t) = g(X_t), \quad (10.26)$$

donde $g(x) = \mathbf{E}h(x + X_{t+h} - X_t)$. La segunda igualdad fue obtenida en el ejemplo 9.6, dado que las variables aleatorias X_t y $X_{t+h} - X_t$ son independientes; por su parte, la primer igualdad se obtiene basándose en la independencia del vector aleatorio $(X_{t_1}, \dots, X_{t_n})$ y la variable aleatoria $X_{t+h} - X_t$.

Dado $x_0 > 0$ decimos que la trayectoria $X_t(\omega)$ alcanza la barrera de nivel x_0 en el intervalo $[0, T]$, si existe $t \in [0, T]$ tal que $X_t(\omega) > x_0$, hecho que equivale a que se verifique $\max_{0 \leq t \leq T} X_t(\omega) > x_0$. Denominamos entonces *problema de barrera con tiempo finito*, al cálculo de la probabilidad

$$\mathbf{P} \left(\max_{0 \leq t \leq T} X_t(\omega) > x_0 \right) = \mathbf{P}(\exists t \in [0, T]: X_t > x_0),$$

es decir, al cálculo de la probabilidad de que el proceso $\{X_t\}$ alcance la barrera de nivel x_0 en el intervalo $[0, T]$. Análogamente se define el *problema de barrera con tiempo infinito*, consistente en el cálculo de la probabilidad

$$\mathbf{P} \left(\max_{t \geq 0} X_t(\omega) > x_0 \right) = \mathbf{P}(\exists t \geq 0: X_t > x_0).$$

Teorema 10.5 (Problema de barrera con tiempo finito).

Consideremos el proceso aleatorio $\{X_t\}$ definido en (10.25). Entonces

$$\mathbf{P} \left(\max_{0 \leq t \leq T} X_t > x_0 \right) = \Phi \left(\frac{-x_0 + aT}{\sqrt{T}} \right) + e^{2ax_0} \Phi \left(\frac{-x_0 - aT}{\sqrt{T}} \right) \quad (10.27)$$

donde $\Phi(x)$ es la distribución normal estándar.

La demostración de este teorema se basa en el teorema 9.3 del muestreo opcional de Doob, mediante el siguiente resultado.

Lema 10.2. Consideremos el proceso aleatorio $\{X_t\}$ definido en (10.25), y las funciones de dos variables

$$A_1(x, t) = e^{2a(x_0-x)} \Phi \left(\frac{x - x_0 - a(T-t)}{\sqrt{T-t}} \right),$$

$$A_2(x, t) = \Phi \left(\frac{x - x_0 + a(T-t)}{\sqrt{T-t}} \right),$$

definidas para x real y t en el intervalo $[0, T)$. Entonces, dados $0 \leq s < t < T$, se verifica

$$\mathbf{E} (A_k(X_t, t) | X_s) = A_k(X_s, s) \quad (k = 1, 2).$$

Demostración. Según hemos visto en el ejemplo 9.6, como las variables aleatorias $Y = X_t - X_s$ y X_s son independientes, se verifica

$$\mathbf{E} (A_k(X_t, t) | X_s) = g_k(X_s) \quad (k = 1, 2),$$

donde $g_k(x) = \mathbf{E} A_k(x + Y, t)$. El lema estará demostrado, si verificamos que $g_k(x) = A_k(x, s)$ ($k = 1, 2$). Designemos $h = t - s$, y $\varphi(x)$ la densidad normal estándar. La variable aleatoria Y tiene distribución normal con parámetros (ah, \sqrt{h}) . Entonces,

$$\begin{aligned} g_1(x) &= \int_{-\infty}^{\infty} e^{2a(x_0 - x - y)} \Phi\left(\frac{x + y - x_0 - a(T - t)}{\sqrt{T - t}}\right) \frac{1}{\sqrt{h}} \varphi\left(\frac{y - ah}{\sqrt{h}}\right) dy \\ &= e^{2a(x_0 - x)} \left[\int_{-\infty}^{\infty} \Phi\left(\frac{x + y - x_0 - a(T - t)}{\sqrt{T - t}}\right) \frac{1}{\sqrt{h}} \varphi\left(\frac{-y - ah}{\sqrt{h}}\right) dy \right] dv \\ &= e^{2a(x_0 - x)} \Phi\left(\frac{x - x_0 - a(T - s)}{\sqrt{T - s}}\right) = A_1(x, s), \end{aligned}$$

donde utilizamos que $e^{-2ay} \varphi((y - ah)/\sqrt{h}) = \varphi((-y - ah)/\sqrt{h})$, y calculamos la integral entre paréntesis rectos de acuerdo a la fórmula (3.33), para la convolución de dos distribuciones, en el caso en que estas son normales. Esto concluye la demostración para el caso $k = 1$. El caso $k = 2$ es análogo (y más sencillo). \square

Demostración del teorema. Consideremos la función $P(x, t) = A_1(x, t) + A_2(x, t)$ (x real, $t \in [0, T)$), donde $A_k(x, t)$ ($k = 1, 2$) son las funciones del lema 10.2. Se verifica

$$P(0, 0) = \Phi\left(\frac{-x_0 + aT}{\sqrt{T}}\right) + e^{2ax_0} \Phi\left(\frac{-x_0 - aT}{\sqrt{T}}\right), \quad (10.28)$$

que es el término a la derecha en (10.27). Para construir una martingala, con valor inicial $P(0, 0)$, consideramos un natural $N \geq 1$, y designamos $t_n = (n/2^N)T$, $F_n = (X_{t_0}, \dots, X_{t_n})$, para cada $n = 0, \dots, 2^N - 1$. En vista del lema 10.2 y de la propiedad de Markov (10.26) obtenemos, que se verifica la propiedad de martingala

$$\mathbf{E} (P(X_{t_{n+1}}, t_{n+1}) | F_n) = P(X_{t_n}, t_n), \quad n = 0, \dots, 2^N - 2.$$

Consideremos el tiempo de parada⁵

$$\tau(N) = \inf\{t_n : X_{t_n} > x_0\} \wedge (1 - 2^{-N})T,$$

y la variable aleatoria

$$\tau = \inf\{t : X_t > x_0, t \in [0, T]\} \wedge T. \quad (10.29)$$

Tomemos límite si $N \rightarrow \infty$. En primer lugar, cuando $\tau(\omega) < T$, tenemos $\tau(N) \rightarrow \tau$, $X_{\tau(N)} \rightarrow X_\tau = x_0$ (dado que las trayectorias son continuas), y, en consecuencia, $P(X_{\tau(N)}, \tau(N)) \rightarrow P(x_0, \tau) = 1$, dado que $P(x_0, t) = 1$, si $t < T$. En segundo lugar, cuando $\tau(\omega) = T$, tenemos $\tau(N) = (1 - 2^{-N})T \rightarrow T$, $X_{\tau(N)} \rightarrow X_T$ (dado que las trayectorias son continuas), y, en este caso, $P(X_{\tau(N)}, \tau(N)) \rightarrow P(X_T, T) = 0$ c.s., dado que $P(x, T) = 0$, si $x < x_0$. En conclusión, obtenemos que

$$P(X_{\tau(N)}, \tau(N)) \rightarrow \mathbf{1}_{\{\tau < T\}} \quad (N \rightarrow \infty) \quad c.s.$$

Además, es sencillo verificar que

$$0 \leq P(X_{\tau(N)}, \tau(N)) \leq 2 + e^{2a(x_0 - X_{t_{2N-1}})},$$

por lo que la sucesión $\{P(X_{\tau(N)}, \tau(N))\}$ es uniformemente integrable. Aplicamos ahora el teorema 9.3 del muestreo opcional de Doob, y tomamos límite si $N \rightarrow \infty$, para obtener

$$P(0, 0) = \mathbf{E} P(X_{\tau(N)}, \tau(N)) \rightarrow \mathbf{E} \mathbf{1}_{\{\tau < T\}} = \mathbf{P} \left(\max_{0 \leq t \leq T} X_t > x_0 \right),$$

lo que, en vista de (10.28), concluye la demostración. \square

Es de particular interés el caso $a = 0$. La fórmula (10.27) en este caso es

$$\mathbf{P} \left(\max_{0 \leq t \leq T} W_t > x_0 \right) = 2\Phi \left(\frac{-x_0}{\sqrt{T}} \right) = 2\mathbf{P}(W_T \geq x_0). \quad (10.30)$$

Una demostración alternativa de la fórmula (10.30) se obtiene mediante el principio de reflexión⁶ a partir del cual resulta, que el proceso $\{V_t\}$ definido mediante

$$V_t(\omega) = \begin{cases} W_t(\omega), & \text{si } s \leq \tau(\omega), \\ 2x_0 - W_t(\omega) & \text{en caso contrario,} \end{cases}$$

⁵Si bien $\tau(N)$ toma los valores $k2^{-N}T$, los resultados del capítulo 9 se extienden sin dificultad a esta situación.

⁶No veremos aquí la demostración de este principio, que se basa en la denominada *propiedad fuerte de Markov*.

es un proceso de Wiener (en el intervalo $[0, T]$), con τ definido en (10.29). Bajo este supuesto

$$\begin{aligned} \mathbf{P}\left(\max_{0 \leq t \leq T} W_t > x_0\right) &= \mathbf{P}\left(\max_{0 \leq t \leq T} W_t > x_0, W_T \geq x\right) \\ &\quad + \mathbf{P}\left(\max_{0 \leq t \leq T} W_t > x_0, W_T < x\right) \\ &= \mathbf{P}(W_T \geq x_0) + \mathbf{P}\left(\max_{0 \leq t \leq T} V_t > x_0, V_T \geq x\right) \\ &= \mathbf{P}(W_T \geq x_0) + \mathbf{P}(V_T \geq x_0) = 2\mathbf{P}(W_T \geq x_0). \end{aligned}$$

Una segunda consecuencia de (10.27) es la resolución del problema de barrera con tiempo infinito. Los sucesos $\{\omega: X_t > x_0 \text{ para algún } t \leq N\}$ ($N = 1, 2, \dots$) forman una sucesión creciente con N , que tiene como límite cuando $N \rightarrow \infty$ al suceso $\{\omega: \exists t \geq 0: X_t > x_0\}$. Tenemos entonces

$$\mathbf{P}(\exists t \geq 0: X_t > x_0) = \lim_{N \rightarrow \infty} \mathbf{P}(\exists t \leq N: X_t > x_0).$$

Para el cálculo del límite anterior, consideremos primero el caso $a \geq 0$. Si $T \rightarrow \infty$ en (10.27), se obtiene

$$\mathbf{P}(\exists t \geq 0: X_t > x_0) = 1.$$

Esto quiere decir que si $a \geq 0$ el proceso $\{X_t\}$ alcanza cualquier barrera con probabilidad uno. El caso $a < 0$ es diferente, y del cálculo del límite si $T \rightarrow \infty$ en (10.27), obtenemos

$$\mathbf{P}\left(\max_{t \geq 0} X_t(\omega) > x_0\right) = \mathbf{P}(\exists t \geq 0: X_t > x_0) = e^{2ax_0}, \quad (10.31)$$

es decir, la variable aleatoria $\max_{t \geq 0} (W_t + at)$ tiene distribución exponencial con parámetro $-2a$.

Consideremos ahora el *problema de dos barreras con tiempo infinito* para el proceso $\{X_t\}$ definido en (10.25), donde $a \neq 0$. Dadas las constantes $y_0 < 0 < x_0$, consideramos las variables aleatorias

$$\tau_1 = \inf\{t \geq 0: X_t > x_0\}, \quad \tau_2 = \inf\{t \geq 0: X_t < y_0\}. \quad (10.32)$$

El problema de dos barreras consiste en determinar $\mathbf{P}(\tau_1 < \tau_2)$, es decir, la probabilidad de alcanzar la barrera positiva antes que la negativa.

(Consideramos el problema análogo para el paseo al azar simple, en los ejemplos 8.5 y 9.16.)

Según hemos visto, si $a > 0$ se tiene $\mathbf{P}(\tau_1 < \infty) = 1$, mientras que si $a < 0$, un razonamiento análogo⁷ conduce a obtener que $\mathbf{P}(\tau_2 < \infty) = 1$. En conclusión, alguna de las dos barreras se alcanza, por lo que se verifica la ecuación

$$\mathbf{P}(\tau_1 < \tau_2) + \mathbf{P}(\tau_1 > \tau_2) = 1. \quad (10.33)$$

Por otra parte, la variable aleatoria $Y = X_{t+h} - X_t$ tiene distribución normal con parámetros (ah, \sqrt{h}) , y la función auxiliar de una variable $P(x) = e^{-2ax}$, verifica la propiedad

$$\mathbf{E} (P(X_{t+h}) - P(X_t) | X_t) = P(X_t) \mathbf{E} (e^{-2aY} - 1) = 0.$$

Es posible verificar que el argumento de aproximación de la demostración del teorema 10.5 se puede aplicar en esta situación. La fórmula que se obtiene, en este caso, es

$$\begin{aligned} 1 = P(0) &= \mathbf{E} P(X_T) = \mathbf{E} e^{-2ax_0} \mathbf{1}_{\{\tau_1 < \tau_2 < T\}} \\ &+ \mathbf{E} e^{-2ay_0} \mathbf{1}_{\{T > \tau_1 > \tau_2\}} + \mathbf{E} P(X_T) \mathbf{1}_{\{T \leq \tau_1 \wedge \tau_2\}}. \end{aligned}$$

Como $P(X_T) \mathbf{1}_{\{T \leq \tau_1 \wedge \tau_2\}} \rightarrow 0$ ($T \rightarrow \infty$) *c.s.* (ya que alguna de las dos barreras se alcanza), y se trata de una variable aleatoria acotada, el último término en la fórmula anterior se anula, si $T \rightarrow \infty$. Al tomar límite si $T \rightarrow \infty$, tenemos

$$1 = e^{-2ax_0} \mathbf{P}(\tau_1 < \tau_2) + e^{-2ay_0} \mathbf{P}(\tau_1 > \tau_2). \quad (10.34)$$

De la resolución del sistema lineal formado por (10.33) y (10.34) se obtiene

$$\mathbf{P}(\tau_1 < \tau_2) = \frac{1 - e^{-2ay_0}}{e^{-2ax_0} - e^{-2ay_0}}. \quad (10.35)$$

que es el resultado buscado⁸. Como consecuencia de este resultado podemos obtener también la fórmula (10.31). En efecto, como el suceso $\{\tau_1 < \tau_2\}$ es creciente con $y_0 \rightarrow -\infty$, se obtiene la identidad (10.31) tomando límite en (10.35), cuando $y_0 \rightarrow -\infty$.

⁷Esta segunda afirmación puede también obtenerse de la primera observando que el proceso $\{-W_t\}$ es un proceso de Wiener.

⁸Comparar con la fórmula (9.21).

10.5. Ejercicios

1. Sea $\{X_t\}$ un proceso aleatorio que parte del origen, con incrementos independientes y estacionarios. Supongamos que la variable aleatoria X_t tiene densidad $p_t(x)$ ($t > 0$). Consideremos los tiempos $0 < t_1 < \dots < t_n < t_{n+1}$, y los reales a_1, \dots, a_n, b , todos arbitrarios. (a) Demostrar la fórmula

$$\begin{aligned} \mathbf{P}(X_{t_1} \leq a_1, \dots, X_{t_n} \leq a_n) \\ = \int_{\{x_1 \leq a_1, \dots, x_1 + \dots + x_n \leq a_n\}} p_{t_1}(x_1) \cdots p_{t_n - t_{n-1}}(x_n) dx_1 \cdots dx_n. \end{aligned}$$

(b) Demostrar que el vector aleatorio $(X_{t_1}, \dots, X_{t_n})$ y la variable aleatoria $Y = X_{t_{n+1}} - X_{t_n}$ son independientes, es decir, se verifica

$$\begin{aligned} \mathbf{P}(X_{t_1} \leq a_1, \dots, X_{t_n} \leq a_n, Y \leq b) \\ = \mathbf{P}(X_{t_1} \leq a_1, \dots, X_{t_n} \leq a_n) \mathbf{P}(Y \leq b). \end{aligned}$$

(c) Demostrar la propiedad de Markov (10.26).

2. Consideremos un proceso aleatorio $\{X_t\}$ que parte del origen, con incrementos independientes y estacionarios. Supongamos que existe $\mathbf{E} X_1$.

(a) Demostrar que existe $\mathbf{E} X_t$, para todo t racional. (b) Demostrar que si existe $\mathbf{E} X_h$, para algún $h > 0$, entonces, $X_{nh}/n \rightarrow \mathbf{E} X_h$ ($n \rightarrow \infty$) *c.s.*

3. Consideremos un proceso aleatorio $\{X_t\}$ con incrementos independientes y estacionarios, tal que existe $\mathbf{E} X_h^2$, para algún $h > 0$. Determinar sucesiones numéricas $\{a_n\}$ y $\{b_n\}$, tales que

$$\mathbf{P}\left(\frac{X_{nh} - a_n}{b_n} \leq x\right) \rightarrow \Phi(x) \quad (n \rightarrow \infty),$$

para todo x real, donde $\Phi(x)$ es la distribución normal estándar.

4. *Ecuación de Cauchy.* Consideremos una función real $f: (0, \infty) \rightarrow \mathbb{R}$, que verifica la *ecuación de Cauchy*:

$$f(x+y) = f(x) + f(y) \text{ para todo } x > 0, y > 0.$$

Demostrar que $f(x) = xf(1)$ para todo $x > 0$, suponiendo que: (a) existe $f'(x)$ ($x > 0$); (b) $f(x)$ es monótona (creciente o decreciente); (c) existe un intervalo, en el que $f(x)$ está acotada. (Sin hipótesis adicionales, el resultado $f(x) = xf(1)$ ($x > 0$) no es cierto.)

5. Consideremos un proceso de conteo $\{N_t\}$, definido como en (10.2), a partir de una sucesión de sumas $\{S_n\}$. Teniendo en cuenta que $N_t \rightarrow \infty$ c.s., demostrar que $S_n \rightarrow \infty$ c.s.

6. Consideremos un proceso de Poisson de parámetro $\alpha > 0$. Dados $0 \leq a_1 < a_2$, demostrar que $\mathbf{P}(S_1 \leq a_1, S_2 \leq a_2) = 1 - e^{-\alpha a_1} - \alpha a_1 e^{-\alpha a_2}$, de dos maneras: (a) a partir de la densidad del vector (S_1, S_2) , (b) a partir de la independencia y la estacionariedad de los incrementos de $\{N_t\}$, y la distribución de N_t . (Esto daría una demostración alternativa del recíproco del teorema 10.1.)

7. Consideremos una sucesión $T_1, U_1, T_2, U_2, \dots$ de variables aleatorias independientes. Supongamos que las variables aleatorias $\{T_n\}$ tienen distribución común exponencial de parámetro $\alpha > 0$; $\{U_n\}$ tienen distribución común exponencial de parámetro $\beta > 0$. Consideremos los procesos de Poisson $\{N_t^\alpha\}$ y $\{N_t^\beta\}$, definidos como en (10.2). Demostrar que el proceso aleatorio $\{N_t^\alpha + N_t^\beta\}$ es un proceso de Poisson, con parámetro que se determinará.

8. Consideremos una sucesión $T_1, X_1, T_2, X_2, \dots$ de variables aleatorias independientes. Supongamos que $\{T_n\}$ tienen distribución común exponencial de parámetro $\alpha > 0$, y que para cada $n = 1, 2, \dots$, se tiene $\mathbf{P}(X_n = 1) = p$, $\mathbf{P}(X_n = 0) = 1 - p$ ($0 < p < 1$). Se define $S_n = X_1 T_1 + \dots + X_n T_n$ ($n = 1, 2, \dots$), y el proceso de conteo $\{N_t\}$ como en (10.2). Demostrar que $\{N_t\}$ es un proceso de Poisson, con parámetro que se determinará.

9. *Transformaciones del proceso de Wiener.* Sea $\{W_t\}$ un proceso de Wiener. Demostrar que los siguientes procesos aleatorios son procesos de Wiener: (a) $\{-W_t\}$; (b) $\{\sqrt{c}W_{t/c}\}$, donde $c > 0$; (c) $\{W_{T+t} - W_T\}$, donde $T > 0$.

10. *Puente browniano.* Sea considera el proceso aleatorio definido mediante

$$B_t = W_t - tW_1 \quad (0 \leq t \leq 1), \quad (10.36)$$

denominado *puente browniano*, donde $\{W_t\}$ es un proceso de Wiener. Demostrar que $\mathbf{P}(B_0 = 0) = \mathbf{P}(B_1 = 0) = 1$, $\mathbf{E} B_t = 0$, $\mathbf{E} B_s B_t = s(1 - t)$ si $0 \leq s \leq t \leq 1$. Demostrar que el proceso $\{B_{1-t}\}_{0 \leq t \leq 1}$ es un puente Browniano, es decir, existe un proceso de Wiener tal que se verifica (10.36).

11. Sean $\{W_t\}$ un proceso de Wiener, y $0 \leq s_1 < t_1 < \dots < s_n < t_n$ tiempos arbitrarios.

(a) Demostrar que $|W_{s_1}|, |W_{t_1} - W_{s_1}|, \dots, |W_{t_n} - W_{s_n}|$ son variables aleatorias independientes.

(b) Demostrar que $W_{s_1}^2, (W_{t_1} - W_{s_1})^2, \dots, (W_{t_n} - W_{s_n})^2$ son variables aleatorias independientes.

12. Sea $\{W_t\}$ un proceso de Wiener. Calcular las siguientes esperanzas condicionales, dado $0 \leq s < t$. (a) $\mathbf{E}(W_t | W_s) = W_s$; (b) $\mathbf{E}(W_t^2 - t | W_s) = W_s^2 - s$.

13. (a) Demostrar que el proceso de Wiener con tendencia $\{X_t\}$ definido en (10.25) tiene incrementos independientes y estacionarios. (b) Calcular la densidad de la variable aleatoria $M_T = \max_{0 \leq t \leq T} X_t$.

14. Consideremos un proceso de Wiener $\{W_t\}$. (a) Demostrar que si $A(x, t) = \Phi((x - x_0)/\sqrt{T - t})$ (x real, $t \in [0, T)$), se verifica la propiedad $\mathbf{E}(A(W_t, t) | W_t) = A(W_s, s)$, donde $0 \leq s < t < T$. (b) Dar una demostración directa de la fórmula (10.30).

15. Verificar la segunda parte del lema 10.2, correspondiente a la función $A_2(x, t)$.

16. Consideremos el proceso aleatorio $\{X_t\}$, donde $X_t = \sigma W_t + at$ ($t \geq 0$), con a y $\sigma > 0$ constantes reales. La constante σ se llama *coeficiente de difusión*. Obtener fórmulas para $\mathbf{P}(\max_{0 \leq t \leq T} X_t(\omega) > x_0)$, y para $\mathbf{P}(\tau_1 < \tau_2)$, para este proceso aleatorio, donde τ_1 y τ_2 se definen en (10.32).

17. Consideremos un proceso de Wiener $\{W_t\}$, y dos constantes $y_0 < 0 < x_0$. Definimos $\tau_1 = \inf\{t \geq 0: W_t > x_0\}$, $\tau_2 = \inf\{t \geq 0: W_t < y_0\}$. Calcular $\mathbf{P}(\tau_1 < \tau_2)$.

Soluciones de algunos ejercicios

Capítulo 1. **1.** **B** acertar en el blanco; **C** acertar en el círculo de radio r_1 ; **D** acertar en el anillo entre los radios r_1 y r_2 . **3.** (a) $\bigcap_{k=1}^n \bar{\mathbf{A}}_k = \overline{\bigcup_{k=1}^n \mathbf{A}_k}$; (b) $\bigcup_{k=1}^n \mathbf{A}_k$; (c) $(\mathbf{A}_1 \bar{\mathbf{A}}_2 \cdots \bar{\mathbf{A}}_n) \cup (\bar{\mathbf{A}}_1 \mathbf{A}_2 \bar{\mathbf{A}}_3 \cdots \bar{\mathbf{A}}_n) \cup \cdots \cup (\bar{\mathbf{A}}_1 \cdots \bar{\mathbf{A}}_{n-1} \mathbf{A}_n)$. **12.** (a) $C_3^4/C_3^9 = 1/21$; (b) $C_3^5/C_3^9 = 5/42$; (c) $C_1^4 C_2^5/C_3^9 = 10/21$. **13.** $1/C_5^{49} = 1/1.906.884 \sim 5,24 \times 10^{-7}$. **15.** (a) $1/9$, (b) $5/18$. **16.** $5/108$. **17.** $(C_{50}^{95} + C_1^5 C_{49}^{95})/C_{50}^{100} \sim 0$, **18.** **18.** $\mathbf{P}(\mathbf{A} \cup \mathbf{B}) = 2/3$, $\mathbf{P}(\mathbf{AB}) = 5/36$, $\mathbf{P}(\mathbf{A}\bar{\mathbf{B}}) = 13/36$, $\mathbf{P}(\bar{\mathbf{A}}\mathbf{B}) = 31/36$. **20.** $(m^K - (m-1)^K)/n^K$. **23.** $11/24$. **24.** (a) $C_3^{15}/C_3^{20} = 91/228$; (b) $(15/20)(14/19)(13/18) = 91/228$. **25.** (a) $0,375$; (b) $0,4$. **26.** $4/25$. **27.** $0,82$; $0,36$. **31.** (a) $0,729$; (b) $0,999$. **32.** $0,9975$. **33.** $1/2$.

Capítulo 2. **1.** (a) $63/64$; (b) $57/64$; (c) $41/64$. **2.** $125/3888 \sim 0,032$. **3.** (a) $0,096$; (b) $0,00427$. **4.** (a) $0,6561$; (b) $0,2916$; (c) $0,3439$. **5.** $p^n(1-p)^n$. **6.** (a) $C_2^5(1/3)^2(2/3)^3 = 80/243 \sim 0,329$; (b) $131/243 \sim 0,539$. **7.** $C_n^{2n-r}(1/2)^{2n-r}$. **8.** (a) $0,384$; (b) $0,992$. **9.** $C_{\ell-1}^{k-1} p^\ell (1-p)^{k-\ell}$. **10.** Tenemos $C_n^{2n}(1/2n)^{2n} \sim 1/\sqrt{\pi n}$ ($n \rightarrow \infty$) por la fórmula de Stirling. **11.** (a) $0,012$; (b) $0,683$. **14.** (a) La probabilidad es menor que 10^{-5} ; (b) $0,98983$; (c) La probabilidad es menor que 3×10^{-4} . **15.** $0,9426$. **16.** $0,9817$. **17.** $0,9596$.

Capítulo 3. **3.** $1/2$. **4.** (a) $0,25$; (b) $0,66$; (c) $0,75$; (d) $0,72$. **5.** (a) $1/2$; (b) $c = 1/2$; (c) $(1 - e^{-1})/2$. **6.** (a) $0,23$; $0,31$; (b) $x = 12,32$. **7.** Resulta $c = 3$; $F(x) = 0$, si $x \leq 0$; $F(x) = x^3$, si $0 < x \leq 1$; $F(x) = 1$, si $x > 1$; $\mathbf{P}(0,1 < X < 0,4) = 0,063$. **10.** $\mathbf{P}(X = 0) = 0,3$; $\mathbf{P}(X = 1) = 0,4$; $\mathbf{P}(X = 2) = 0,3$. **12.** (a) La constante $c = 3/2$; (d) Tenemos $p_X(x) = x/2$ si $0 \leq x \leq 2$; $p_X(x) = 0$ si $x < 0$, ó si $x > 2$; $p_Y(y) = 3y^2$ si $0 \leq y \leq 2$; $p_Y(y) = 0$ si $y < 0$ ó si $y > 2$. **13.** Sí, no. **15.** (a) $c = 2$; (b) Sí. **17.** $F(x) = 0$ para $x < 3$; $F(x) = 1/20$ para $3 \leq x < 4$; $F(x) = 1/5$ para $4 \leq x < 5$;

$F(x) = 1/2$ para $5 \leq x < 6$; $F(x) = 1$ para $x \geq 6$; $\mathbf{P}(X \geq 5) = 4/5$.

18. Se tiene $\mathbf{P}(Y \leq x) = F((x-b)/a)$; $p_Y(x) = (1/a)p((x-b)/a)$. **23.**

$\mathbf{P}(Y \leq x) = F(x^{1/3})$; $\mathbf{P}(Z \leq x) = 1 - F(x-)$. **26.** $p_{X+Y} = x$, si $0 \leq x \leq 1$; $p_{X+Y} = 2 - x$, si $1 < x \leq 2$; $p_{X+Y} = 0$, si $x < 0$ ó $x > 2$.

Capítulo 4. 1. $(a+b)/2$; $(b-a)^2/12$; $(a^2+b^2)(a+b)/4$. **3.** $\mathbf{E}(X^k) = 0$ si k

es impar; $\mathbf{E}(X^k) = 1 \cdot 3 \cdot 5 \cdots (k-1) = (k-1)!!$ si k es par; $\mathbf{E}(e^{hX}) = e^{h^2/2}$

para todo $h \in \mathbb{R}$. **4.** $1/\alpha$; $1/\alpha^2$; $6/\alpha^3$. **12.** $\mathbf{E}X = a$; $\mathbf{var} X = 2b^2$. **13.**

Tenemos $\mathbf{E}X = 2\alpha/\sqrt{\pi}$ y $\mathbf{var} X = (3/2 - 4/\pi)\alpha^2$. **15.** (a) $\mathbf{E}X = (n+1)/2$;

$\mathbf{var} X = (n^2 - 1)/12$; (b) $\mathbf{E}X = n$; $\mathbf{var} X = n(n-1)$. **22.** $\exp(\lambda e^h - 1)$.

23. $(q + pe^h)^n$. **24.** $\ln 2$. **26.** $11/9$; $5/9$. **28.** (a) $\mathbf{E}Z = 10$; $\mathbf{var} Z = 13$; (b)

$\mathbf{E}Z = 10$; $\mathbf{var} Z = 13$; (c) $\mathbf{E}Z = 10$; $\mathbf{var} Z = 15, 4$. **29.** (a) $1 - r^2$; (b) 0.

Capítulo 5. 13. Sí. **15.** Sí. **17.** No.

Capítulo 6. 1. (a) $\text{sen}(lt)/(lt)$; (b) $f(t) = 1 - |t|$, si $|t| \leq 1$; $f(t) = 0$, si

$|t| > 1$. **5.** (a) No; (b) Sí; (c) Sí; (d) No; (e) Sí; (f) Sí; (g) No. **6.** $9/2$. **18.**

(b) (i) Sí; (ii) Sí; (iii) No; (iv) No; (v) No.

Capítulo 7. 1. $1 - \Phi(0, 78) = 0, 22$. **2.** $1 - \Phi(1, 31) = 0, 095$.

Capítulo 8. 6. Las potencias valen

$$\mathbf{P}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}, \quad \mathbf{P}^3 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$\mathbf{P}^{3n+1} = \mathbf{P}$, $\mathbf{P}^{3n+2} = \mathbf{P}^2$, $\mathbf{P}^{3n+3} = \mathbf{P}^3$. Todos los estados tienen período 3.

10. (a) $q = p_0 + p_1q + p_2q^2 + p_3q^3 + \cdots$. (b) $q = 1/2$. **17.** $\pi = u$. **18.** Si n es la cantidad de puntos de la región, tenemos $\pi = (1/n, \dots, 1/n)$.

Capítulo 9. 3. (b) Sí. **10.** (b) $a_n = \sum_{k=1}^n \ln \mathbf{E} e^{X_k}$. **11.** (b) El compensador

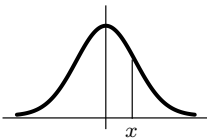
$A_n = \sum_{k=1}^n \mathbf{E} X_k^2$. **13.** (a) Sí; (b) No; (c) Sí; (d) Sí; (e) Sí. **17.** (b) No necesari-

amente. **19.** El recíproco no es cierto en ningún caso: (a), (c), (d).

20. (a) Sí; (b) Sí.

Capítulo 10. 17. $|y_0|/(x_0 + |y_0|)$.

Tabla de la densidad normal estándar

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


	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3555	0.3538
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2637	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2155	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0893	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0025	0.0025
3.2	0.0024	0.0023	0.0022	0.0022	0.0021	0.0020	0.0020	0.0019	0.0018	0.0018
3.3	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0013
3.4	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010	0.0009	0.0009

Bibliografía

- [1] A.A. Borovkov, *Probability theory*, Gordon and Breach: New York 1998.
- [2] W. Feller, *Introducción a la teoría de probabilidades y sus aplicaciones*, Editorial Limusa: Mexico D.F., vol. I: 1983, vol II: 1978.
- [3] B.V. Gnedenko, *Theory of probability*, Chelsea, Springer: New York 1962.
- [4] A. Gut, *An intermediate course in probability*, Springer: New York 1995.
- [5] V.V. Petrov, *Limit theorems of probability theory*, Oxford University Press: Oxford, 1995.
- [6] S. M. Ross, *Introduction to probability models*. Harcourt/Academic Press: San Diego, CA, 2000.
- [7] A. N. Shiryaev, *Probability*, Springer: New York, 1996.

