

PROBABILIDADES

José E. Valdés Castro

PROBABILIDADES

JOSÉ E. VALDÉS CASTRO



FACULTAD DE MATEMÁTICA Y COMPUTACIÓN
UNIVERSIDAD DE LA HABANA

Copyright © 2023 José E. Valdés Castro

ESTAS NOTAS CONSTITUYEN EL TEXTO CORRESPONDIENTE A LA ASIGNATURA TEORÍA DE LAS PROBABILIDADES DE LA CARRERA MATEMÁTICA EN LA UNIVERSIDAD DE LA HABANA. ACTUALMENTE ES UN TRABAJO EN PROGRESO. NO DUDE EN CONTACTAR AL AUTOR SI DESEA HACER CUALQUIER SUGERENCIA SOBRE SU CONTENIDO.

Probability . . . we must carefully distinguish three aspects of the theory:

a) the formal logic content,

b) the intuitive background,

c) the applications.

The character, and the charm, of the whole structure cannot be appreciated without considering all three aspects in their proper relation.

William Feller

(An Introduction to Probability Theory and Its Applications, Volume I)

ÍNDICE GENERAL

Índice general	V
Prefacio	IX
I Probabilidad y sucesos	1
1 Probabilidad	2
1.1. Introducción	2
1.2. Probabilidad de un suceso	3
1.3. Definición clásica	5
1.4. Método geométrico	6
1.5. Método estadístico	10
1.6. Teoría de las probabilidades y estadística matemática	15
1.7. Problemas y ejercicios	16
2 Simulación estocástica	18
2.1. Simulación	18
2.2. Problema de la ruina del jugador	20
2.3. Método de Monte Carlo	22
2.4. Problemas y ejercicios	24
3 Elementos de combinatoria	25
3.1. Preliminares	25
3.2. Selección con reposición	27
3.3. Selección sin reposición	28
3.4. Problemas y ejercicios	33
4 Sobre decisión estadística	35
4.1. Un problema de mercadeo	35

4.2. Un ejemplo sobre control de calidad	37
4.3. Problemas y ejercicios	38
5 Definición de probabilidad y propiedades	40
5.1. Sucesos y sigma-álgebra	40
5.2. Definición de probabilidad	44
5.3. Propiedades básicas de la probabilidad	46
5.4. Modelo discreto	49
5.5. Información contenida en un espacio medible	51
5.6. Continuidad de la probabilidad	53
5.7. Lema de Borel-Cantelli. Parte I	57
5.8. Problemas y ejercicios	58
6 Probabilidad condicional e independencia de sucesos	60
6.1. Probabilidad condicional	60
6.2. Probabilidad total y fórmula de Bayes	63
6.3. Independencia de sucesos	67
6.4. Lema de Borel-Cantelli. Parte II	71
6.5. Problemas y ejercicios	73
7 Espacio producto	75
7.1. Espacio producto	75
7.2. Modelo de Bernoulli	77
7.3. Modelo multinomial	82
7.4. Sucesión infinita de experimentos independientes	83
7.5. Método del primer paso	86
7.6. Probabilidad de extinción de una población	87
7.7. Problemas y ejercicios	89
8 Complementos	92
8.1. Conjunto no medible	92
8.2. Sobre un espacio muestral infinito no numerable	94
8.3. Infinitos experimentos independientes	95
II Variables aleatorias	98
9 Variable aleatoria discreta	99
9.1. Variable aleatoria. Distribución de probabilidad	99
9.2. Valor esperado	109
9.3. Sobre estadística	115

9.4. Simulación	117
9.5. Noción sobre la teoría de juegos	122
9.6. Problemas y ejercicios	123
10 Variable aleatoria en general	127
10.1. Introducción	127
10.2. Sigma-álgebra de Borel	128
10.3. Variable aleatoria	132
10.4. Distribución de probabilidad	134
10.5. Independencia de variables aleatorias	136
10.6. Operaciones con variables aleatorias	137
10.7. Problemas y ejercicios	143
11 Valor Esperado	145
11.1. Variable aleatoria simple	145
11.2. Variable aleatoria no negativa	149
11.3. Variable aleatoria arbitraria	155
11.4. Fórmula de cambio de variable	158
12 Densidad de probabilidad	160
12.1. Integrales de Riemann y de Lebesgue	160
12.2. Densidad de probabilidad	162
12.3. Varianza	166
12.4. Distribución normal	169
12.5. Distribuciones Gamma, Chi-cuadrado y Beta	171
12.6. Covarianza y coeficiente de correlación	172
12.7. Problemas y ejercicios	176
13 Función de distribución	179
13.1. Función de distribución	179
13.2. Variable aleatoria mixta	185
13.3. Sobre estadística	186
13.4. Método de Monte Carlo	189
13.5. Funciones de una variable aleatoria	190
13.6. Problemas y ejercicios	192
14 Complementos	194
14.1. Identidad de Wald	194
14.2. Espacio L^2 de variables aleatorias	197

III Vectores aleatorios	201
15 Vectores Aleatorios	202
15.1. Vectores aleatorios	202
15.2. Distribución de vectores aleatorios	204
15.3. Vector de variables aleatorias discretas	208
15.4. Vector de variables aleatorias con densidad de probabilidad	210
15.5. Simulación de variables aleatorias con distribución normal	216
15.6. Distribución condicional. Esperanza condicional	218
15.7. Distribución conjunta normal	224
15.8. Sobre estadística	227
15.9. Problemas y ejercicios	230
IV Función característica. Teoremas límites	234
16 Función característica	235
16.1. Función característica	235
16.2. Función característica de la distribución normal	241
16.3. Función característica de la distribución Gamma	242
16.4. Función característica y momentos	243
16.5. Problemas y ejercicios	245
17 Convergencias. Teoremas Límites	246
17.1. Desigualdad de Chebyshev	246
17.2. Ley de los grandes números	247
17.3. Convergencia en distribución	256
17.4. Teorema Central del Límite	264
17.5. Problemas y ejercicios	266
18 Complementos	268
18.1. Teorema de Rényi	268
V Nota histórica	271

PREFACIO

Estas notas constituyen una introducción a la teoría de las probabilidades, aunque en ellas se consideran también, de forma intuitiva, algunas ideas que conciernen a la estadística.

El cálculo de probabilidades se inició en el siglo XVII vinculado a los juegos de azar. En la actualidad la teoría de las probabilidades y la estadística matemática se aplican en numerosos problemas de la física, la ingeniería, la economía, la biología, las ciencias sociales y la medicina, por ejemplo: en el diseño de equipos y sistemas con alta confiabilidad y en la planificación de sus mantenimientos; en las decisiones de mercadeo; en el control de la calidad de productos; en la organización de redes de comunicación y de computadoras; en la organización de procesos productivos; en los test médicos y ensayos clínicos, etc.

Los primeros cuatro capítulos de este texto constituyen una introducción al cálculo de probabilidades, la simulación, y también a algunas nociones sobre estadística. La teoría de las probabilidades como tal se expone a partir del Capítulo 5.

La mayoría de los ejercicios que se proponen en el interior del texto complementan la teoría y deben ser hechos por los estudiantes.

En algunos ejemplos, y en algunos ejercicios y problemas propuestos al final de cada capítulo de la Parte I, las soluciones se apoyan en ideas intuitivas; pero la intuición puede ser engañosa, por lo tanto debe usarse con cuidado.

Varios problemas y ejercicios propuestos al final de los capítulos, que podrían tener un alto grado de dificultad en sus soluciones, se señalan con el símbolo *. A veces se brindan las respuestas de los ejercicios.

En la elaboración de este texto ha sido imprescindible la colaboración de la profesora Marelys Crespo Navas.

Parte I

Probabilidad y sucesos

CAPÍTULO 1

PROBABILIDAD

En este capítulo se examinan tres métodos de asignación de probabilidades a los eventos, ampliamente usados en los cálculos de probabilidades.

1.1. Introducción

Los tres métodos de asignación de probabilidades que se examinarán tienen gran importancia, pues en muchas ocasiones posibilitan la aplicación de la teoría, y facilitan además el desarrollo de la intuición, muy útil en esta área de la matemática. Estos métodos no forman parte de la teoría moderna de las probabilidades, cuyo desarrollo tiene una base axiomática y se inició en el año 1933 por el matemático ruso A. N. Kolmogorov, quien para ello se apoyó en los resultados de la teoría de probabilidades elaborada hasta ese entonces, conjuntamente con los resultados de otras áreas de la matemática que iniciaron su desarrollo a principios del siglo XX. La teoría moderna de las probabilidades se comenzará a exponer a partir del Capítulo 5.

En la Parte V de este texto, Nota histórica, se presenta un breve resumen del desarrollo histórico de la teoría de las probabilidades. Señalemos ahora solo cuatro momentos de gran importancia en este desarrollo: la solución de varios problemas de juegos de azar contenidos en cartas escritas en el año 1654 por B. Pascal y P. Fermat, que forman parte de la correspondencia entre estos dos científicos; la publicación en el año 1713 de *El Arte de las Conjetu-*

ras, obra de Jacob Bernoulli, que puede considerarse el primer gran aporte al desarrollo de la teoría; la obra *Teoría Analítica de las Probabilidades* de P. S. Laplace, editada por primera vez en el año 1812, en la cual se exponen de manera sistemática los resultados obtenidos hasta esa época y se presentan otros nuevos; y la publicación en 1933 de *Fundamentos de la Teoría de las Probabilidades*, de A. N. Kolmogorov, donde este autor, como ya se mencionó, construye la teoría de las probabilidades sobre una base axiomática.

1.2. Probabilidad de un suceso

Ejemplo 1.1 Consideremos un experimento que consiste en el lanzamiento al azar de un dado que tiene sus caras numeradas desde el número 1 hasta el 6. Atribuiremos al término *azar* el significado intuitivo que usualmente se le otorga. En este experimento son posibles seis resultados que se excluyen mutuamente y, si se considera que el dado está equilibrado, o sea, es simétrico y homogéneo, lo cual por supuesto es una idealización, entonces estos seis resultados son igualmente posibles. El resultado del experimento es *aleatorio* (casual), pues depende del azar, y por lo tanto impredecible. Sin embargo, a través de la *probabilidad* puede medirse la posibilidad de que en el experimento ocurra algún suceso (evento) de interés.

¿Cuál es la probabilidad del suceso que consiste en que al lanzar el dado aparezca un número par? Una posible respuesta intuitiva a esta pregunta es que la probabilidad de que ocurra ese suceso es igual a $1/2$. En efecto, una medida razonable de la posibilidad de aparición del suceso de interés es la proporción del número de casos favorables a su ocurrencia, con respecto al total de posibles resultados del experimento. Puesto que la aparición de cualquiera de los números 2, 4 ó 6 lleva implícita la aparición de un número par, entonces del total de resultados, tres son favorables a la aparición de este suceso; por lo tanto, concluimos que la probabilidad de aparición de un número par es $3/6 = 1/2$. De manera similar se calcula la probabilidad de que al lanzar el dado aparezca un número mayor que 4, esta es, $2/6 = 1/3$.



El experimento del lanzamiento al azar de un dado constituye un ejemplo muy simple de **experimento aleatorio**. Un experimento o fenómeno se dice que es aleatorio si sus resultados dependen del azar, es decir, son casuales, y, por lo tanto, desconocidos antes de su realización. Un hecho o resultado

de interés que puede ocurrir o no en un experimento aleatorio se denomina **suceso o evento**.

Experimentos o fenómenos aleatorios de mayor complejidad y utilidad que el lanzamiento al azar de un dado son: el tiempo de vida de un organismo vivo, o el tiempo hasta el fallo de un equipo; el tiempo de servicio, reparación o procesamiento de un objeto por una persona o dispositivo; el precio de los productos un día determinado en el mercado internacional; el movimiento Browniano de partículas.

*Desde el punto de vista de las aplicaciones, la **probabilidad** de un suceso es una medida de la posibilidad de ocurrencia de este suceso.*

El concepto de probabilidad está presente en numerosos problemas de la actividad humana, por ejemplo: cuando se diseña un equipo, es de interés estimar la probabilidad de que, en un periodo de tiempo dado, este pueda cumplir sin fallar las funciones para las cuales se diseñó; en el tema de seguros de vida es importante estimar la probabilidad de que una persona viva, digamos, más de 60 años, dado que actualmente tiene 24 años; en el diseño de un centro de servicios tiene interés determinar el número de unidades de servicio que garantiza que, con una probabilidad dada, el tiempo de espera de un cliente no supere un valor prefijado.

Los resultados de un experimento aleatorio se identifican con los elementos de cierto conjunto denominado **espacio muestral**, y denotado por el símbolo Ω . A cada elemento $\omega \in \Omega$ se le llama **punto muestral**. En el experimento del lanzamiento del dado, el conjunto Ω está constituido por seis puntos muestrales correspondientes a la aparición de cada una de las caras del dado, por ejemplo, $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Muchas veces en las aplicaciones el interés no consiste en el resultado concreto ω de un experimento, sino en si este resultado pertenece a uno u otro subconjunto de Ω . Por eso, desde el punto de vista práctico, los sucesos son subconjuntos del espacio muestral Ω para los cuales, a partir de la realización del experimento, es posible decidir sobre su ocurrencia o no, es decir, es posible decidir si el punto muestral pertenece o no a un subconjunto dado. Los sucesos se denotan por las letras mayúsculas A, B, C , etc., sin subíndices o con subíndices, y se identifican con subconjuntos de Ω .

En el ejemplo del lanzamiento del dado, el suceso que consiste en la aparición de un número par se identifica con el subconjunto $A = \{2, 4, 6\}$ y el que consiste en la aparición de un número mayor que 4 con el subconjunto $B = \{5, 6\}$.

1.3. Definición clásica

El método de asignación de probabilidades a los sucesos utilizado en el Ejemplo 1.1, fue formalizado por P. Laplace a comienzos del siglo XIX, y constituyó la primera definición de probabilidad. Este método recibe el nombre de **definición clásica** de probabilidad y puede formularse en general de la siguiente manera:

Se considera un experimento con un conjunto finito Ω de resultados aleatorios, mutuamente excluyentes e *igualmente posibles* (equiprobables). Denotemos por A un suceso relacionado con estos resultados. La probabilidad del suceso A , denotada por $P(A)$, se calcula por la fórmula

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

donde $|\Omega|$ denota el número total de resultados posibles del experimento, y $|A|$ el número de estos resultados favorables a la ocurrencia del suceso A , es decir, que conllevan la ocurrencia de A .

La igualdad de posibilidades de los resultados es una noción intuitiva, que podría asumirse o no de acuerdo a las condiciones específicas del problema que se esté examinando; por ejemplo, podría asumirse debido a condiciones como simetría u homogeneidad en el experimento o fenómeno.

Note que según el método de la definición clásica, la probabilidad de cualquier suceso A satisface $0 \leq P(A) \leq 1$, en particular, $P(\Omega) = 1$. Para el conjunto vacío se toma $P(\emptyset) = 0$.

En el experimento del lanzamiento del dado, el suceso que consiste en la aparición de un número menor que 8 se identifica con el conjunto Ω , luego su probabilidad es 1; el suceso que consiste en la aparición de un número mayor que 9 se identifica con el conjunto vacío, por lo tanto la probabilidad de este suceso es 0.

Al espacio muestral Ω se le llama **suceso seguro** o **suceso cierto**. Puesto que contiene todos los resultados del experimento, este suceso siempre ocurre. El conjunto \emptyset no contiene resultados del experimento y se denomina **suceso imposible**.

Sean ahora $A \subset \Omega$ y $A^c = \Omega \setminus A$. Usando la definición clásica de probabilidad se tiene que $P(A^c) = 1 - P(A)$, o de forma equivalente, $P(A) = 1 - P(A^c)$.

Efectivamente,

$$P(A^c) = \frac{|A^c|}{|\Omega|} = \frac{|\Omega \setminus A|}{|\Omega|} = \frac{|\Omega| - |A|}{|\Omega|} = 1 - P(A).$$

Al conjunto A^c , que representa la no ocurrencia de A , se le llama **suceso contrario** del suceso A .

Ejemplo 1.2 Se realiza el lanzamiento al azar de dos dados equilibrados. ¿Cuál es la probabilidad de que en ambos dados aparezca el seis? ¿Cuál es la probabilidad de que la suma de los puntos de los dos dados sea igual a 3? ¿Cuál es la probabilidad de que en al menos uno de los dados aparezca el seis?

El lanzamiento de los dos dados constituye un experimento compuesto por dos experimentos, cada uno de los cuales corresponde al lanzamiento de uno de los dados, y cuyos conjuntos de posibles resultados Ω_1 y Ω_2 se representan por el mismo conjunto $\Omega_1 = \Omega_2 = \{1, 2, 3, 4, 5, 6\}$. Al experimento compuesto se le asocian los resultados del conjunto $\Omega = \Omega_1 \times \Omega_2$.

El suceso que consiste en la aparición del seis en ambos dados se asocia con el conjunto $A = \{(6, 6)\}$, y el suceso que consiste en que la suma de los puntos es 3 con el conjunto $B = \{(1, 2), (2, 1)\}$. Entonces por la definición clásica se tiene que

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1}{36} \quad \text{y} \quad P(B) = \frac{|B|}{|\Omega|} = \frac{2}{36} = \frac{1}{18}.$$

El conjunto $D = \{1, 2, \dots, 5\} \times \{1, 2, \dots, 5\}$ se asocia con el suceso que consiste en que en ninguno de los dados aparece el 6, contrario al suceso que consiste en la aparición de al menos un 6. Luego la probabilidad del suceso de interés D^c se calcula,

$$P(D^c) = 1 - P(D) = 1 - \frac{25}{36} = \frac{11}{36}.$$

■

1.4. Método geométrico

Ejemplo 1.3 Imaginemos que una persona llega en un instante aleatorio a una parada de ómnibus, y que los ómnibus de la ruta de su interés llegan consecutivamente a la parada exactamente cada 20 minutos. ¿Cuál es la probabilidad de que la persona tenga que esperar más de 5 minutos para viajar?

El intervalo $\Omega = [0, 20]$ puede considerarse el conjunto de los posibles instantes de llegada de la persona a la parada, donde los minutos 0 y 20 representan los arribos de un ómnibus y del siguiente, respectivamente. El conjunto Ω es el espacio muestral. Por otro lado, el tiempo de espera es mayor que 5 minutos cuando la persona llega en el subintervalo $A = (0, 15)$, luego A constituye el conjunto de resultados asociados a la ocurrencia del suceso para el cual se desea calcular la probabilidad.

Tanto Ω como A son conjuntos infinitos, y por consiguiente no se puede utilizar la definición clásica para el cálculo. Sin embargo, es posible utilizar una idea análoga a la de la definición clásica, asignando como probabilidad del suceso el cociente entre las longitudes de A y de Ω . En este caso la probabilidad es igual a $15/20 = 0,75$. ■

El método de asignación de una probabilidad utilizado en el ejemplo anterior se conoce con el nombre de **método geométrico** (o **probabilidad geométrica**). A continuación se formula el método de manera general.

Sean Ω un subconjunto de \mathbb{R}^n que tiene volumen (longitud, área, etc), finito y mayor que 0, y que constituye el total de resultados del experimento aleatorio, y A una región de Ω que también tiene volumen, constituida por los resultados asociados a la ocurrencia de un suceso de interés.

Si se puede suponer que la probabilidad de ocurrencia del suceso no depende de la ubicación y forma de la región A , sino solo de su volumen, entonces la probabilidad de A , denotada por $P(A)$, se calcula por la fórmula

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)},$$

donde $\lambda(C)$ denota el volumen de un subconjunto C de \mathbb{R}^n .

Observe que, similar a la definición clásica, se tiene: $0 \leq P(A) \leq 1$, $P(\Omega) = 1$ y $P(A) = 1 - P(A^c)$. Se toma $P(\emptyset) = 0$.

El símbolo λ se utilizará indistintamente para denotar, longitud, área, etc. Según el contexto del problema específico tratado, quedará claro a qué se refiere este símbolo.

Es obvio que en los modelos asociados con la probabilidad geométrica, los conjuntos abiertos, y sus correspondientes cerrados que incluyen las fronteras, proporcionan el mismo resultado en los cálculos.

Note, intuitivamente, que de forma análoga a la definición clásica, con la probabilidad geométrica se asume que todos los puntos de Ω tienen “igual

posibilidad" de ser seleccionados (aunque en realidad la probabilidad de seleccionar un punto determinado es igual a 0). Tampoco se requiere para el cálculo la realización de experimento alguno, pero igualmente se suponen condiciones de homogeneidad y, además, la finitud del volumen del espacio muestral Ω .

En lo adelante, las frases "punto aleatorio" y "selección de un punto al azar" de una región de \mathbb{R}^n , significarán que se asumen las condiciones requeridas para el cálculo de la probabilidad geométrica.

En los problemas que requieren el cálculo de una probabilidad geométrica es conveniente auxiliarse de figuras en las cuales se representen las regiones cuyos volúmenes se desea calcular.

Ejemplo 1.4 *Problema del encuentro.* Dos personas acuerdan encontrarse en un sitio dado entre dos instantes de tiempo, digamos, 0 y T ; asumamos que con seguridad ambas personas llegarán en el intervalo $[0, T]$. El primero que llegue espera por el otro durante un tiempo τ , $0 < \tau < T$, y en el caso que el segundo no llegue durante ese tiempo, se marcha. Se asume que los instantes de llegadas de las dos personas son aleatorios y constituyen dos experimentos *independientes*¹. ¿Cuál es la probabilidad de que las dos personas no se encuentren?

Se tiene un experimento compuesto por dos experimentos, cuyos resultados, en cada caso, son los puntos del intervalo $[0, T] = \Omega_i$, $i = 1, 2$, que representan los instantes en que llegan las personas. Los resultados del experimento compuesto son los puntos del conjunto $\Omega = \Omega_1 \times \Omega_2 = [0, T] \times [0, T]$. El suceso que consiste en el encuentro de las dos personas se identifica con el conjunto

$$A = \{(x, y) \in \Omega : |x - y| \leq \tau\}.$$

Este conjunto se representa por la región sombreada en el diagrama de la Figura 1.1, y tiene área $\lambda(A) = T^2 - (T - \tau)^2$. De acuerdo con la probabilidad geométrica, se tiene que

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{T^2 - (T - \tau)^2}{T^2}.$$

La probabilidad de que no se produzca el encuentro es igual a

$$1 - P(A) = \left(1 - \frac{\tau}{T}\right)^2.$$

¹El concepto de experimentos independientes intuitivamente significa que los resultados de un experimento no influyen en los resultados de cualquier otro experimento.

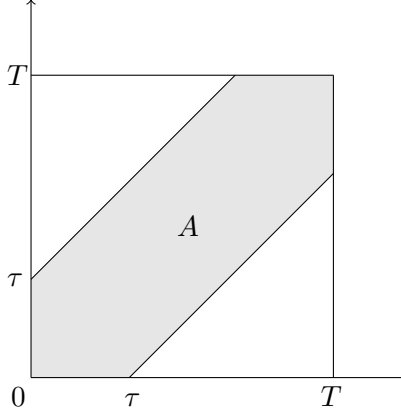


Figura 1.1: Diagrama para el Problema del encuentro.

Ejemplo 1.5 Dos puntos seleccionados al azar del intervalo $[0, l]$ lo dividen en tres partes. ¿Cuál es la probabilidad de que las tres partes sean menores o iguales que $l/2$?

Denotemos por x y y los puntos seleccionados del intervalo $[0, l]$. El conjunto Ω de posibles resultados del experimento puede representarse,

$$\Omega = \{(x, y) : 0 \leq x \leq l, 0 \leq y \leq l\}.$$

El suceso para el cual deseamos calcular su probabilidad se identifica con el subconjunto $A = A_1 \cup A_2$ de Ω , donde

$$A_1 = \{(x, y) \in \Omega : x < y, x \leq l/2, y - x \leq l/2, l - y \leq l/2\}$$

$$A_2 = \{(x, y) \in \Omega : y \leq x, y \leq l/2, x - y \leq l/2, l - x \leq l/2\}.$$

El conjunto A se representa por la región sombreada en el diagrama de la Figura 1.2. No es difícil hallar que $\lambda(A_1) = \lambda(A_2) = l^2/8$. Entonces, usando la probabilidad geométrica, se tiene que

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{2(l^2/8)}{l^2} = \frac{1}{4}.$$

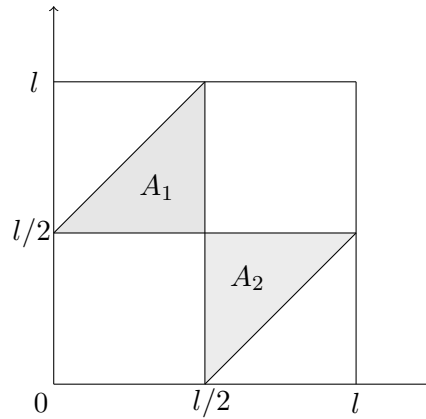


Figura 1.2: Diagrama para el Ejemplo 1.5.

1.5. Método estadístico

En muchos fenómenos no están presentes las condiciones requeridas para el uso de la definición clásica o la probabilidad geométrica. Esta es la situación que se presenta cuando una moneda o un dado no están equilibrados, y se desea calcular la probabilidad de que al lanzarlos ocurra determinado suceso. También, en la observación del tiempo de vida de un equipo, donde puede considerarse como espacio muestral el intervalo $\Omega = [0, \infty)$ y el subconjunto de Ω , $A = (a, \infty)$, $a > 0$, se asocia con el suceso que consiste en que el fallo del equipo ocurre después del instante a , es evidente que para asignar una probabilidad a A no pueden ser usadas ni la definición clásica ni la probabilidad geométrica. ¿Cómo calcular en casos como estos, al menos aproximadamente, la probabilidad de un suceso?

Pudiera parecer que en los experimentos y fenómenos aleatorios no existen leyes, y por lo tanto es imposible realizar predicciones en estos fenómenos. Sin embargo, esto no es así, la experiencia acumulada de muchos años demuestra que en numerosas ocasiones la aleatoriedad se somete a leyes, a regularidades, las llamadas *leyes o regularidades estadísticas*.

Si realizamos una gran cantidad de lanzamientos al azar de una moneda, observaremos que a medida que el número de lanzamientos crece, la **frecuencia relativa** de aparición de una cara dada de la moneda, es decir, la proporción m/n del número m de apariciones de esta cara con respecto al total n de lanzamientos, tiende a estabilizarse alrededor de cierto valor. Cuando la moneda es equilibrada (lo cual es una idealización) este valor es igual a $1/2$.

En la Figura 1.3 se presentan los gráficos correspondientes a las frecuencias relativas de aparición de una cara de una moneda equilibrada, para varias series de mil lanzamientos, obtenidas por medio de la simulación (vea el Capítulo 2). En el eje de las abscisas se representa el número de lanzamientos y en el eje de las ordenadas la frecuencia relativa correspondiente, además los puntos obtenidos se unieron por segmentos rectilíneos. El gráfico correspondiente a una de las series se dibujó en negro, y los correspondientes a las restantes series en gris.

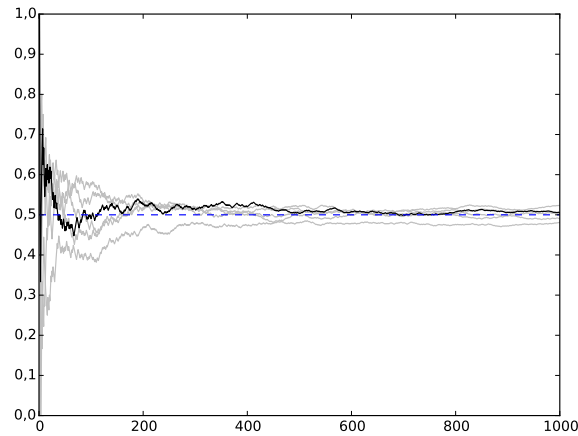


Figura 1.3: Estabilización de la frecuencia relativa.

Asumamos que un experimento o fenómeno aleatorio es observado reiteradamente, de manera tal que se obtiene una serie de experimentos iguales e independientes. Supongamos que en cada experimento puede ocurrir o no un suceso A , y que se realizan k series de estos experimentos. Denotemos por n_i y m_i , respectivamente, el número total de experimentos y el número de experimentos en los cuales ocurre A en la serie i , $i = 1, 2, \dots, k$. Resulta que cuando n_i es grande, independientemente de los resultados individuales obtenidos en los experimentos de cada serie, las frecuencias relativas m_i/n_i , $i = 1, 2, \dots, k$, de ocurrencia del suceso A en cada una de las k series de experimentos, son aproximadamente iguales entre sí, es decir,

$$\frac{m_1}{n_1} \approx \frac{m_2}{n_2} \approx \dots \approx \frac{m_k}{n_k}.$$

Además, estas frecuencias relativas se estabilizan alrededor de cierto número, digamos p , en general desconocido, con una tendencia a aproximarse a

este número a medida que la cantidad de experimentos de las series crece. El número p caracteriza la posibilidad de ocurrencia de A .

Esta propiedad de estabilización de la frecuencia relativa de ocurrencia de A alrededor de un número p , en una serie de experimentos idénticos e independientes, establecida empíricamente, constituye el ejemplo más simple de ley estadística, y permite considerar a p como el valor de la probabilidad del suceso A .

Luego, la probabilidad $P(A)$ de un suceso A se puede calcular aproximadamente por la fórmula

$$p = P(A) \approx \frac{m}{n},$$

donde m es el número de experimentos en los cuales ocurrió A en una serie de un número grande n de experimentos idénticos e independientes.

A esta forma de asignación de un valor de probabilidad a un suceso se le llama **método estadístico** o método de la frecuencia relativa.

Note que una probabilidad $P(A)$, calculada aproximadamente usando la frecuencia relativa, también cumple, $0 \leq P(A) \leq 1$, $P(\Omega) = 1$ y $P(A^c) = 1 - P(A)$.

Una dificultad del método estadístico consiste en que, a diferencia de la definición clásica y la probabilidad geométrica, para usar este método se requiere la realización reiterada de un número grande n de experimentos independientes y en condiciones idénticas. Además, si se realizan varias series de experimentos, incluso con igual número de experimentos en cada una de ellas, las frecuencias relativas correspondientes a cada serie serán en general diferentes entre sí.

Posteriormente se verá que p es el límite de la frecuencia relativa m/n cuando n tiende a infinito, en ciertos sentidos del límite que serán definidos.

La frecuencia relativa es importante en la *interpretación estadística* de las probabilidades que se obtienen en los diferentes cálculos. Por ejemplo, la probabilidad $1/3$ de que aparezca un número mayor que 4 al lanzar al azar un dado equilibrado se puede interpretar de la siguiente manera. Si el dado se lanzara un número grande de veces, digamos 300 veces, entonces, puesto que $1/3 = 100/300$, se puede concluir que en aproximadamente 100 de esas ocasiones aparecería un número mayor que 4. En el Ejemplo 1.3 sobre la espera de una persona por un ómnibus, la probabilidad 0,75 se puede interpretar co-

mo que de 100 veces que la persona llegue a la parada, en aproximadamente 75 de ellas tendrá que esperar más de 5 minutos.

El siguiente ejemplo ilustra uno de los problemas de la estadística, que consiste en **estimar**, es decir, hallar aproximadamente, a partir de los datos observados en una **muestra** de una **población**, alguna característica de esa población.

Ejemplo 1.6 En un embalse hay una cantidad N desconocida de peces. ¿Cómo calcular aproximadamente el valor de N ?

Se capturan M peces, se marcan de alguna manera, y se devuelven al embalse. Después de esperar a que los peces marcados se mezclen con los restantes, entonces del total de peces del embalse, los cuales constituyen la *población*, se captura al azar una *muestra* de n peces. La característica de interés es el tamaño N de la población. Denotemos por m el número de peces marcados de la muestra.

La probabilidad de que al seleccionar al azar un pez del embalse este sea uno de los marcados, es igual a $\frac{M}{N}$ según la definición clásica de probabilidad. Por otra parte, según el método estadístico, esta probabilidad es aproximadamente igual a la proporción $\frac{m}{n}$ de peces marcados de la muestra, cuando el valor de n es grande. Entonces tiene lugar la aproximación:

$$\frac{M}{N} \approx \frac{m}{n}, \quad (1.1)$$

de donde

$$N \approx M \frac{n}{m}.$$

Si, por ejemplo, se marcaron $M = 100$ peces de la presa, y el tamaño de la muestra seleccionada al azar es $n = 200$, de los cuales $m = 5$ peces resultaron estar marcados, entonces la cantidad aproximada de peces en la presa es 4000.

En general, el método descrito puede ser utilizado para calcular aproximadamente el total de integrantes de especies de animales que habitan en determinadas regiones.

Supongamos ahora que se desea estimar la cantidad desconocida M de artículos defectuosos que contiene un lote grande de N artículos, pero sin inspeccionar todos los artículos del lote. Aquí la población está constituida por los N artículos del lote, y la característica de interés es M . Si se toma una muestra al azar de n artículos del lote y m de ellos resultan defectuosos,

entonces usando de nuevo la fórmula 1.1 se obtiene,

$$M \approx N \frac{m}{n}.$$

■

Ejemplo 1.7 *Ley estadística.* Un experimento aleatorio que generaliza el experimento del lanzamiento al azar de una moneda, y en el cual también se observa fácilmente la presencia de una ley estadística, consiste en la realización de una serie de n experimentos, n grande, en cada uno de los cuales se lanza al azar un número k , por ejemplo $k = 5$, de monedas idénticas. Cuando lanzamos 5 monedas al azar pueden aparecer 0, 1, 2, 3, 4 ó 5 escudos. Al realizar una serie de estos experimentos, para $n = 1000$, se obtuvieron los siguientes resultados: en 50 experimentos aparecieron 0 escudos; en 152 apareció 1 escudo; en 291 aparecieron 2 escudos; en 308 aparecieron 3 escudos; en 157 aparecieron 4 escudos, y en 42 aparecieron 5 escudos. Las frecuencias relativas correspondientes son: 0,05, 0,152, 0,291, 0,308, 0,157 y 0,042, respectivamente.

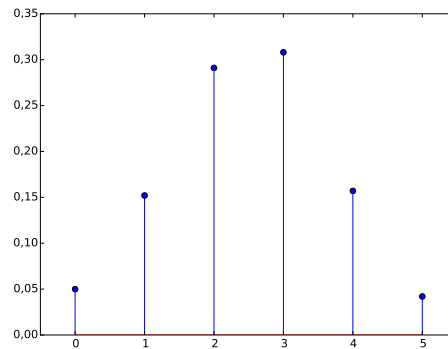


Figura 1.4: Cinco monedas. Frecuencias relativas del número de escudos

Si se dibuja un gráfico en un sistema de coordenadas cartesianas, donde en el eje de las abscisas se coloca el número de escudos que pueden aparecer al lanzar las cinco monedas, y en el eje de las ordenadas se coloca la frecuencia relativa correspondiente a estos valores en la serie de 1000 experimentos (vea la Figura 1.4), observaremos que los puntos del gráfico se distribuyen de forma “acampanada”. Esta forma se observará siempre que se repitan series con un número grande de experimentos, y con monedas aproximadamente

equilibradas, independientemente de los resultados particulares que se obtienen en cada uno de los lanzamientos.

Aquí la ley estadística consiste en la distribución de los puntos en forma acampanada. A medida que la probabilidad de aparición de una de las caras sea significativamente mayor que la de la otra cara, la distribución de los puntos será más asimétrica. ■

Como se mencionó al inicio de este capítulo, los tres métodos examinados de asignación de valores numéricos a las probabilidades de los sucesos, es decir, la definición clásica, la probabilidad geométrica y el método estadístico, no forman parte de la teoría moderna de las probabilidades, pues esta se desarrolla de manera axiomática; sin embargo, estos tres métodos tienen una importancia enorme en las aplicaciones y en la interpretación de muchos resultados teóricos.

1.6. Teoría de las probabilidades y estadística matemática

La teoría de las probabilidades y la estadística matemática son dos áreas de la matemática muy relacionadas, pero que tienen objetos de estudio diferentes. En la primera se asume un modelo para un fenómeno o experimento aleatorio, se deducen propiedades y fórmulas para el modelo, y realizando cálculos adecuados se predicen, en términos probabilísticos, resultados que se obtendrían en el experimento aun sin observar este. Una parte importante de la segunda es la inferencia estadística, la cual, a partir de los datos observados (muestra) en el experimento, indica cómo inferir el modelo que los describe y estimar parámetros de interés.

El Ejemplo 1.2 sobre el lanzamiento de dos dados es un problema de probabilidades. En este se asume el modelo que permite usar la definición clásica de probabilidad y se calculan probabilidades para predecir posibles resultados, pero sin realizar experimentos. Si en este mismo problema los dados se asumen no equilibrados, entonces surge un problema de la estadística, que consiste en la estimación de las probabilidades utilizando el método de la frecuencia relativa. El Ejemplo 1.6, sobre el embalse de peces y el lote de artículos, aunque puede considerarse como un problema de ambas áreas, estadística y probabilidades, esencialmente es un problema de estadística, pues se trata de la estimación de los parámetros N en un caso, y M en el otro caso, a partir de los datos observados.

Resumamos algunos rasgos distintivos de estas dos áreas de la matemática:

La *teoría de las probabilidades* proporciona modelos para el estudio de las leyes estadísticas, y métodos para la obtención de probabilidades desconocidas de sucesos aleatorios de interés a partir de las probabilidades conocidas de otros sucesos aleatorios.

La *estadística matemática*, usando los datos de observaciones en fenómenos y experimentos aleatorios, proporciona métodos para la selección adecuada de los modelos probabilísticos y la realización de inferencias sobre estos fenómenos.

1.7. Problemas y ejercicios

1. Un dado equilibrado se lanza n veces al azar. ¿Cuál es la probabilidad de que el número 6 aparezca al menos una vez? ¿cuál es el mínimo valor de n para que esta probabilidad sea mayor que 0,9?
2. Se realiza la planificación de un torneo con tres etapas, eliminatorias, semifinales y final, en el cual participarán 8 boxeadores; luego, mediante un sorteo, se aparean los boxeadores. En cada tope entre dos boxeadores se elimina al perdedor. En el torneo participan dos boxeadores que se sabe que ganarán con seguridad sus topes con los seis restantes. ¿Cuál es la probabilidad de que estos dos boxeadores ocupen el primero y el segundo lugar en el torneo? **R:** 4/7.
3. **Juego justo con una moneda no equilibrada.* Considere dos jugadores que desean realizar un juego equitativo mediante el lanzamiento de una moneda al azar, pero la moneda que poseen no es equilibrada. Justifique, intuitivamente, por qué con el siguiente algoritmo la probabilidad de aparición de cada cara es igual a 1/2:
 1. Lanzar la moneda.
 2. Lanzar la moneda otra vez.
 3. Si en ambos lanzamientos aparece la misma cara, reiniciar con el paso 1.
 4. Si en los dos lanzamientos aparecen caras diferentes, el resultado del último lanzamiento es el deseado.

4. Considere un cuadrado de lado 1. Se selecciona un punto aleatorio del cuadrado. ¿Cuál es la probabilidad de que el punto se encuentre a una distancia menor que $1/2$ del centro del cuadrado?
5. Después de finalizar su trabajo Carlos llega todos los días a una parada de dos ómnibus en cualquier instante entre las 5:00 y las 6:00 p.m. Las rutas 1 y 2 llegan a la parada de acuerdo con los siguientes horarios:

ruta 1 : 5:15 5:22 5:32 5:50 6:06

ruta 2 : 5:16 5:25 5:33 5:52 6:10

Carlos siempre coge el primer ómnibus que llega a la parada. Si este es de la ruta 2, pasa por casa de Elvira, y si es de la ruta 1, pasa por casa de Carmen. Elvira se queja de que en 3 meses Carlos la ha visitado sólo 8 veces. ¿Cómo explica usted esta situación? Calcule la probabilidad p de que Carlos tome la ruta 2. Considere que en los tres meses hay 80 días de trabajo. **R:** $p \approx 0,117$.

6. En las condiciones del *Problema del encuentro*, Ejemplo 1.4, suponga que las dos personas, Carlos y Elvira, esperan una por la otra hasta el encuentro, ¿cuál es la probabilidad de que Carlos tenga que esperar por Elvira un tiempo mayor que t ? ¿Cuál es la probabilidad de que no tenga que esperar?
7. *Se seleccionan dos puntos al azar, x y y , del intervalo $(0, l)$ ¿Cuál es la probabilidad de que con los segmentos de longitudes x , y y l pueda formarse un triángulo? **R:** $1/2$.
8. *Dos puntos seleccionados al azar del intervalo $(0, l)$ lo dividen en tres partes. ¿Cuál es la probabilidad de que las tres partes formen un triángulo? **R:** $1/4$.

SIMULACIÓN ESTOCÁSTICA

La **simulación** consiste en la representación aproximada de procesos y sistemas del mundo real por medio de otros procesos y sistemas artificiales, pero análogos, que resultan más asequibles para su estudio. Está vinculada con la generación, observación y análisis de una historia de estos procesos y sistemas artificiales, con el propósito de hacer inferencias acerca de los reales.

Si el fenómeno investigado es aleatorio y transcurre en el tiempo, como puede ser el desarrollo durante un tiempo de un proceso o sistema, entonces habitualmente se usa el término *estocástico* como equivalente del término aleatorio y se hace referencia a un proceso o sistema estocástico. Los modelos utilizados para el estudio de los fenómenos aleatorios son probabilísticos y la simulación de estos se denomina **simulación estocástica**.

2.1. Simulación

La realización de cualquier estudio de simulación estocástica está basada en la posibilidad de generar números del intervalo $(0, 1)$ con la computadora, aproximadamente con igual probabilidad. Estos números constituyen una sucesión finita y habitualmente son obtenidos mediante un algoritmo, que se inicia con un número llamado *semilla*. Por lo tanto, en realidad no son números aleatorios, razón por la cual reciben el nombre de *números pseudoaleatorios*.

En general, los algoritmos que se utilizan para generar los números pseu-

doaleatorios simulan muy bien los aleatorios. Actualmente todos los lenguajes de programación, y los software de computación que lo requieren, tienen incorporadas instrucciones para la generación de números pseudoaleatorios, cuya ejecución está asociada a las palabras *random*, *randomize*, *uniform*, etc. En algunos casos el usuario selecciona la semilla que iniciará la generación de una sucesión de estos números, en otros casos la semilla es seleccionada por el propio programa, por ejemplo, utilizando el reloj de la computadora.

A continuación, de manera intuitiva, se exponen algunas nociones básicas sobre simulación estocástica.

Examinemos cómo simular la ocurrencia de un suceso A , tal que $P(A) = p$. Sea \tilde{A} el suceso que consiste en que un punto aleatorio del intervalo $(0, 1)$, denotado por U , es menor que p , y representemos por $(U < p)$ este suceso, o sea, $\tilde{A} = (U < p)$. Si se asume el modelo de la probabilidad geométrica, se obtiene $P(\tilde{A}) = p$. Los sucesos A y \tilde{A} tienen igual probabilidad p , de ahí que la ocurrencia de \tilde{A} simula la ocurrencia de A .

Se genera ahora con la computadora un número aleatorio (en realidad pseudoaleatorio) del intervalo $(0, 1)$, el cual se toma como valor de U , y si este número es menor que p (suceso \tilde{A}), se considera que ocurrió A . En el lenguaje Python un número aleatorio del intervalo $(0, 1)$ se puede generar con la función *uniform* $(0, 1)$.

Si se desea calcular la frecuencia relativa de ocurrencia de un suceso A en una serie de n experimentos idénticos, entonces se repite n veces la generación de números aleatorios del intervalo $(0, 1)$ y se cuenta el número de veces m que ocurre \tilde{A} (y por lo tanto A). Con el crecimiento de n , la frecuencia relativa m/n tiende a aproximarse a la probabilidad p .

Ejemplo 2.1 Con el programa en lenguaje Python que se presenta en la Figura 2.1, se simula una serie de n lanzamientos de una moneda con probabilidad de aparición de escudo en cada lanzamiento igual a p . Para la ejecución del programa se tomaron los valores $p = 1/2$ y $n = 1000$. Este programa fue usado para la simulación de las series de sucesiones de lanzamientos de una moneda equilibrada, cuyos gráficos asociados se mostraron en la Figura 1.3.

■

Habitualmente la simulación estocástica se utiliza para calcular aproximadamente alguna probabilidad u otra característica desconocida de interés. Con el programa de la Figura 2.1 el objetivo consiste en simular los lanzamientos reiterados de una moneda, solo para mostrar cómo la frecuencia relativa de aparición del escudo se aproxima a la probabilidad conocida p . En la

próxima sección, a través de un problema particular, se muestra cómo usando la simulación es posible calcular aproximadamente una probabilidad que podría ser desconocida.

```
from random import uniform
n=1000 # número de lanzamientos
p=1/2  # probabilidad de aparición de escudo
i=0    # número de orden de los lanzamientos
m=0    # número acumulado de escudos hasta i
print("frec_relativa_en_el_lanzamiento_i:")
print("i", "frec")
while i<n:
    i+=1
    u = uniform(0,1)
    if u < p:
        m+=1
    frec=m/i
    print(i, round(frec,3))
```

Figura 2.1: Simulación de una serie de lanzamientos de una moneda.

2.2. Problema de la ruina del jugador

Dos jugadores, digamos A y B, inicialmente poseen un capital de a y b unidades, respectivamente. Asumiremos que a y b son números naturales. Cuando un jugador gana una jugada, recibe una unidad del capital de su contrario. La probabilidad de que B gane una jugada dada es p , $0 < p < 1$, y la probabilidad de que gane A es $1 - p$. El juego (partida) finaliza cuando uno de los jugadores se arruina, es decir, en el momento en que uno de los jugadores acumula un capital igual a $a + b$, y en que, por supuesto, el capital del contrario es igual a 0. Puede demostrarse que siempre se produce la ruina de alguno de los jugadores. Se desea hallar la probabilidad p_b de que B arruine a A.

El programa en lenguaje Python de la Figura 2.2 describe el algoritmo de simulación para el cálculo aproximado de p_b . Para la ejecución del programa se tomaron los valores, $a = 15$, $b = 5$, $p = 4/7$ y el número de partidas $n = 1000$.

Sea q_a la probabilidad de que A arruine a B. Note que $p_b + q_a = 1$. Analice-

```

from random import uniform
a=15      # capital inicial de A
b=5       # capital inicial de B
p=4/7     # probabilidad de que B gane una jugada
n=1000    # número de repeticiones (partidas)
m=0       # número de partidas ganadas por B de las n
c=a+b
for i in range(n):
    B=b # capital de B en cada jugada de una partida
    while(0<B<c):
        u = uniform(0,1)
        if u<p:
            B+=1
        else:
            B-=1
    if B>0:
        m+=1
p_b=m/n # valor aproximado de p_b
print(p_b)

```

Figura 2.2: Simulación del Problema de la ruina del jugador.

mos el caso $p = 1/2$. En este caso, intuitivamente, podemos hacer la conjetura de que p_b y q_a son proporcionales a b y a , respectivamente, es decir, $p_b = c_1 b$ y $q_a = c_2 a$, donde $0 < c_1 < 1/b$ y $0 < c_2 < 1/a$. Suponiendo cierta la conjetura, se tiene que $c_1 b + c_2 a = 1$. Como esta ecuación se satisface para todos los valores de a y b , si y solo si $c_1 = c_2$, entonces $c_1 = c_2 = 1/(a + b)$, de donde,

$$p_b = \frac{b}{a+b} \quad \text{y} \quad q_a = \frac{a}{a+b}.$$

Tomando en el programa de simulación, $p = 1/2$, y diferentes juegos de valores de a y b , la ejecución del programa puede sugerir la validez o no de la conjetura hecha: si en la ejecución del programa se obtienen valores aproximados de p_b muy diferentes de $b/(a + b)$, esto sugiere que la conjetura podría no ser cierta, y si los valores son muy cercanos, que la conjetura podría ser verdadera; en el segundo caso tenemos razones para continuar el análisis teórico del problema con el objetivo de probar la conjetura.

2.3. Método de Monte Carlo

Las técnicas de simulación estocástica se usan, además, en la solución de problemas de naturaleza no estocástica, por ejemplo, en el cálculo aproximado de integrales definidas y de volúmenes de regiones. A estas técnicas para la solución de problemas no estocásticos se les denomina en conjunto **método de Monte Carlo**.

Ilustremos un método sencillo de cálculo aproximado de integrales definidas (existen otros métodos más eficientes).

Ejemplo 2.2 Sea la función $g(x)$ integrable según Riemann en el intervalo $[0, 1]$, tal que $0 \leq g(x) \leq 1$, $x \in [0, 1]$. Se desea calcular aproximadamente la integral

$$I = \int_0^1 g(x) dx.$$

Consideremos $\Omega = [0, 1] \times [0, 1]$. Denotemos la región bajo la curva de $g(x)$ y sobre el intervalo $[0, 1]$ por A , o sea

$$A = \{(x, y) \in \Omega : y < g(x)\}.$$

El valor de I coincide con el área de A . Sean U_1 y U_2 dos puntos aleatorios del intervalo $(0, 1)$ y, por consiguiente (U_1, U_2) un punto aleatorio de Ω . Note que la probabilidad del suceso A es igual a la probabilidad $P((U_1, U_2) \in A)$, de que el punto aleatorio (U_1, U_2) pertenezca a la región A . Por el método de la probabilidad geométrica se halla que

$$P((U_1, U_2) \in A) = P(A) = \frac{\lambda(A)}{\lambda(\Omega)} = \lambda(A) = I.$$

(Recordemos que $\lambda(A)$ denota el área de A). Es decir, $I = P(A)$. Por otra parte, por el método de la frecuencia relativa se tiene, cuando n es grande, que

$$P(A) \approx \frac{m}{n},$$

donde n es el total de puntos aleatorios de la región Ω generados y m el número de estos puntos que pertenecen a la región A . Entonces

$$I = P(A) \approx \frac{m}{n},$$

cuando n es grande. En conclusión, para calcular la integral aproximadamente, se genera con la computadora un número grande n de puntos aleatorios

del cuadrado $[0, 1] \times [0, 1]$ y se cuenta el número de estos puntos m que pertenecen a la región A . El cociente m/n es el valor aproximado de la integral.

■

Observe ahora, intuitivamente, que un número aleatorio del intervalo (a, b) se genera por la fórmula $(b - a)U + a$, donde U es un número aleatorio del intervalo $(0, 1)$. Entonces un punto aleatorio (X, Y) del rectángulo $(a, b) \times (c, d)$ se genera usando las fórmulas

$$X = (b - a)U_1 + a \quad \text{y} \quad Y = (d - c)U_2 + c,$$

donde U_1 y U_2 son números aleatorios del intervalo $(0, 1)$. El siguiente ejemplo muestra cómo la idea contenida en el ejemplo anterior puede ser generalizada.

Ejemplo 2.3 Sea la función integrable $g : [a, b] \rightarrow [c, d]$, donde $c \geq 0$. El área de la región

$$A = \{(x, y) \in [a, b] \times [c, d] : y < g(x)\},$$

se puede calcular aproximadamente por la fórmula

$$\lambda(A) \approx (b - a)(d - c) \frac{m}{n},$$

donde n (grande), es el total de puntos seleccionados al azar del rectángulo $(a, b) \times (c, d)$, y m el número de estos puntos que pertenecen a la región A . Note que si $c = 0$, entonces

$$\lambda(A) = \int_a^b g(x) dx.$$

■

La técnica aquí expuesta para el cálculo de integrales definidas se extiende de manera natural para el cálculo aproximado de integrales múltiples.

Observación 1 Los métodos numéricos tradicionales de cálculo aproximado de integrales definidas de funciones de una variable son muy eficientes. Sin embargo, en el caso de integrales múltiples de un orden elevado el método de Monte Carlo podría ser de gran utilidad.

2.4. Problemas y ejercicios

1. Implemente un programa de simulación en lenguaje Python para el *Problema del encuentro*, Ejemplo 1.4, que permita calcular aproximadamente la probabilidad de encuentro cuando $T = 60$ y $\tau = 20$.
2. Implemente un programa de simulación en lenguaje Python para el Problema del *Juego justo con una moneda no equilibrada*, Ejercicio 3 de la Sección 1.7, y verifique que la frecuencia relativa de aparición de cada una de las caras es aproximadamente igual a $1/2$.
3. Sea $\Omega = \{1, 2, \dots, N\}$. Explique, intuitivamente, por qué con la expresión $\lfloor UN \rfloor + 1$, donde U es un punto aleatorio del intervalo $(0, 1)$, se puede simular la selección aleatoria de los puntos de Ω cada uno con probabilidad $1/N$. Por $\lfloor x \rfloor$ se denota la parte entera del número real x . Demuestre que $P(\lfloor UN \rfloor + 1 = k) = 1/N, 1 \leq k \leq N$.
4. Considere un círculo inscrito en un cuadrado. Implemente un programa en lenguaje Python para calcular aproximadamente el valor de π usando el método de Monte Carlo.
5. Usando el método de Monte Carlo calcule aproximadamente la integral

$$\int_0^2 x(1 - x/2) dx.$$

Compare este resultado con el valor exacto de la integral, obtenido por métodos analíticos.

6. *Formule cómo calcular aproximadamente, usando el método de Monte Carlo, la integral de una función $g(x, y)$ en una región D . Calcule aproximadamente la integral

$$\iint_{x^2+y^2 \leq R^2} e^{-(x^2+y^2)} dx dy,$$

donde, por ejemplo, $R = 2$. Compare este resultado con el valor exacto de la integral, obtenido por métodos analíticos.

ELEMENTOS DE COMBINATORIA

La teoría combinatoria se utiliza en diversas áreas de la matemática y en la solución de numerosos problemas prácticos. En este capítulo se considerarán las nociones más básicas de combinatoria, expuestas desde el punto de vista de su aplicación en los cálculos de probabilidades.

3.1. Preliminares

Sea el conjunto Ω_i tal $|\Omega_i| = n_i, i = 1, 2, \dots, k$, donde $|D|$ denota el número de elementos de un conjunto D finito. Consideremos el experimento que consiste en la selección de k elementos, uno y solo uno de cada conjunto Ω_i .

La selección de un elemento del conjunto $\Omega_i, i = 1, 2, \dots, k$, es también un experimento que llamaremos **prueba**. Entonces el experimento inicial está compuesto de una sucesión de k pruebas. Asociado con este experimento el conjunto de resultados es $\Omega = \Omega_1 \times \dots \times \Omega_k$. Note que $|\Omega| = n_1 n_2 \dots n_k$.

Ejemplo 3.1 Sean dos bolsas numeradas 1 y 2. La bolsa 1 contiene 2 bolas rojas y 3 negras, y la bolsa 2 contiene 3 bolas rojas y 4 negras. Se selecciona al azar una bola de cada bolsa. ¿Cuál es la probabilidad del suceso A que consiste en que las dos bolas seleccionadas sean negras?

El experimento está compuesto por dos pruebas sucesivas, que consisten en las selecciones de una bola de la bolsa 1 y una bola de la bolsa 2. Nume-

remos las bolas de las bolsas 1 y 2 con los números naturales desde 1 hasta 5 y desde 1 hasta 7, respectivamente. Los símbolos R_{ij} y N_{ij} indicarán que una bola es roja, respectivamente negra, y que pertenece a la bolsa i y tiene número de orden j . Los conjuntos Ω_1 y Ω_2 que contienen los posibles resultados de la selección en las bolsas 1 y 2, respectivamente, pueden representarse como

$$\Omega_1 = \{R_{11}, R_{12}, N_{13}, N_{14}, N_{15}\} \quad \text{y} \quad \Omega_2 = \{R_{21}, R_{22}, R_{23}, N_{24}, N_{25}, N_{26}, N_{27}\}.$$

El conjunto de resultados del experimento es $\Omega = \Omega_1 \times \Omega_2$ y, por las condiciones del problema, estos resultados se asumen igualmente probables. El suceso de interés se expresa como el subconjunto de Ω ,

$$A = \{N_{13}, N_{14}, N_{15}\} \times \{N_{24}, N_{25}, N_{26}, N_{27}\}.$$

Como $|\Omega| = 5 \times 7 = 35$ y $|A| = 3 \times 4 = 12$, utilizando el método clásico de cálculo de una probabilidad obtenemos,

$$P(A) = 12/35 \approx 0,34.$$

Por lo tanto, si el experimento se realiza, por ejemplo, 100 veces, entonces en aproximadamente 34 de las ocasiones las dos bolas seleccionadas serán negras. ■

Sea ahora, $\Omega_i = \{0, 1\}$, $i = 1, 2, \dots, k$. Entonces

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_k = \{(x_1, x_2, \dots, x_k) : x_i \in \{0, 1\}, i = 1, \dots, k\}$$

y $|\Omega| = 2^k$. En este caso Ω se puede interpretar como el conjunto de resultados de un experimento constituido por una sucesión de k pruebas, en cada una de las cuales puede ocurrir o no cierto suceso de interés. A la ocurrencia del suceso le llamamos *éxito* y le asociamos el número 1, y a su no ocurrencia le llamamos *fracaso* y le asociamos el número 0.

Ejemplo 3.2 Una moneda equilibrada se lanza tres veces al azar. ¿Cuál es la probabilidad del suceso A que consiste en que en el segundo lanzamiento aparece el escudo? ¿Cuál es la probabilidad del suceso B que consiste en que el escudo aparece al menos dos veces?

El experimento está constituido por una sucesión de tres pruebas (lanzamientos). Asociemos la aparición del escudo con un éxito, y la de la otra cara con un fracaso.

Consideremos ternas ordenadas con componentes 0 y 1, donde la primera componente de cada terna se asocia al resultado correspondiente al primer lanzamiento, su segunda componente al resultado correspondiente al segundo lanzamiento y su tercera componente al resultado del tercero. Sea $\Omega_i = \{0, 1\}$ el conjunto de resultados del lanzamiento i , $i = 1, 2, 3$. El conjunto de resultados correspondiente a los tres lanzamientos es

$$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3 = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \{0, 1\}\}.$$

Además,

$$A = \{(1, 1, 1), (1, 1, 0), (0, 1, 1), (0, 1, 0)\}.$$

$$B = \{(1, 1, 1), (1, 1, 0), (0, 1, 1), (1, 0, 1)\}.$$

Note que $|\Omega| = 2^3 = 8$. Puesto que la moneda es equilibrada, cada elemento de Ω tiene igual probabilidad. Usando el método clásico, se halla que

$$P(A) = P(B) = 4/8 = 1/2.$$

■

3.2. Selección con reposición

Sean $\Omega_i = L$, $i = 1, 2, \dots, k$, $|L| = n$ y $\Omega = \Omega_1 \times \dots \times \Omega_k$. Entonces $|\Omega| = n^k$. Ahora Ω se puede interpretar como el conjunto de resultados correspondiente a un experimento de **selección con reposición**, que consiste en seleccionar sucesivamente k elementos del conjunto L , los cuales se repone en este conjunto después de cada selección y antes de la siguiente. Note que cada selección constituye en sí un experimento que ya antes le hemos llamado prueba.

Ejemplo 3.3 Una bolsa contiene 3 bolas rojas y 2 negras. Se seleccionan sucesivamente tres bolas al azar, reponiendo después de cada selección la bola antes de la próxima elección. ¿Cuál es la probabilidad de que de las tres bolas seleccionadas, 2 sean rojas y 1 negra?

Numeremos las bolas del número 1 al 5 y denotemos las bolas de forma similar a como se hizo en el Ejemplo 3.1. El conjunto de posibles resultados en la selección i se puede representar por el conjunto

$$\Omega_i = \{R_1, R_2, R_3, N_4, N_5\}, \quad i = 1, 2, 3.$$

Entonces el conjunto de resultados del experimento es

$$\Omega = \Omega_1 \times \Omega_2 \times \Omega_3 \quad \text{y} \quad |\Omega| = 5^3 = 125.$$

Sea A el suceso de interés. El suceso A puede ocurrir de tres formas diferentes: o en las dos primeras elecciones las bolas son rojas y en la tercera es negra; o en la primera y tercera elección las bolas son rojas y en la segunda es negra; o en la segunda y tercera elección las bolas son rojas y en la primera es negra. A cada una de estas tres formas le corresponden $3 \times 3 \times 2$ resultados. Por consiguiente, $|A| = 3 \times (3 \times 3 \times 2) = 54$, de donde,

$$P(A) = 54/125 = 0,432.$$

■

3.3. Selección sin reposición

Variaciones

Sea un conjunto L y $|L| = n$. Supongamos que el experimento consiste en la **selección sin reposición** de k elementos de L , $1 \leq k \leq L$, lo cual significa que después de cada selección el elemento seleccionado no se repone en L . El conjunto Ω de resultados del experimento lo constituyen todas las formas posibles de seleccionar k elementos de entre el total de n . A cada una de estas formas se le llama **variación de n en k** y su número, $|\Omega|$, se denota por V_n^k .

Veamos cómo se calcula V_n^k . Numeremos los elementos de L con los números naturales desde 1 hasta n . En la primera selección están disponibles los n elementos; la segunda selección se realiza entre los restantes $n - 1$ elementos; para la tercera selección quedan disponibles $n - 2$ elementos; en general, la selección número k se realiza entre los $n - (k - 1)$ elementos que quedan disponibles. Entonces

$$V_n^k = n(n - 1) \cdots (n - (k - 1)) = \frac{n!}{(n - k)!}.$$

En particular, el número de variaciones de n en n , $V_n^n = n!$, es el número de **permutaciones** de n elementos.

Ejemplo 3.4 En la siguiente tabla se representan las variaciones de 4 en 3 cuando $L = \{1, 2, 3, 4\}$.

123	132	213	231	312	321
124	142	214	241	412	421
134	143	314	341	413	431
234	243	324	342	423	432

Si ahora $L = \{1, 2, 3\}$, entonces las permutaciones de los elementos de L son:

$$123 \quad 132 \quad 213 \quad 231 \quad 312 \quad 321.$$

■

Combinaciones

Sea de nuevo un conjunto L y $|L| = n$. Cada subconjunto de L constituido por k elementos, $1 \leq k \leq n$, se llama **combinación de n en k** , y el número de estas combinaciones se denota por $\binom{n}{k}$. Definimos, $\binom{n}{0} = 1$ (el conjunto vacío \emptyset es un subconjunto de L que no contiene ninguno de sus elementos).

Note que el número de combinaciones de n en k y el número de combinaciones de n en $n - k$ coinciden, es decir, $\binom{n}{k} = \binom{n}{n-k}$.

Una combinación de n en k es un conjunto de elementos de tamaño k , donde no importa el orden de los elementos. Por el contrario, para determinar las variaciones de n en k se tienen en cuenta todos los órdenes posibles entre los elementos de cada combinación. Luego a cada combinación de k elementos, corresponden $k!$ variaciones, de donde

$$V_n^k = k! \binom{n}{k}.$$

Por lo tanto, el número de combinaciones se calcula por la fórmula

$$\binom{n}{k} = \frac{V_n^k}{k!} = \frac{n!}{(n-k)!k!}.$$

Ejemplo 3.5 Si $L = \{1, 2, 3, 4\}$, entonces las combinaciones de 4 en 3 correspondientes a L son los $\binom{4}{3} = 4$ subconjuntos,

$$\{1, 2, 3\} \quad \{1, 2, 4\} \quad \{1, 3, 4\} \quad \{2, 3, 4\}.$$

■

Ejemplo 3.6 Consideremos un experimento que consiste en la realización de n pruebas, en cada una de las cuales puede ocurrir éxito o fracaso. ¿De cuántas formas pueden ocurrir k éxitos? Si suponemos que el experimento consiste en el lanzamiento al azar 5 veces de una moneda equilibrada, ¿cuál es la probabilidad de que el escudo aparezca exactamente 3 veces?, ¿cuál es la probabilidad de que aparezca al menos una vez?

El conjunto Ω de los resultados correspondientes a este experimento se puede representar de la siguiente manera

$$\Omega = \{(x_1, x_2, \dots, x_n) : x_i \in \{0, 1\}, i = 1, \dots, n\}.$$

Entonces el conjunto A , constituido por todas las formas posibles de obtener k éxitos, se expresa

$$A = \left\{ (x_1, x_2, \dots, x_n) \in \Omega : \sum_{i=1}^n x_i = k \right\}.$$

Por consiguiente, exactamente k éxitos pueden ocurrir de $|A| = \binom{n}{k}$ formas.

Note que $|\Omega| = 2^n$. Puesto que la moneda está equilibrada, todos los puntos de Ω tienen igual probabilidad, y entonces la probabilidad $p_n(k)$ de que el escudo aparezca exactamente k veces en n lanzamientos se calcula por la fórmula

$$p_n(k) = \frac{\binom{n}{k}}{2^n}.$$

En particular, la probabilidad de que el escudo aparezca 3 veces en los 5 lanzamientos se calcula

$$p_5(3) = \frac{\binom{5}{3}}{2^5} = 5/16 = 0,3125.$$

Si la aparición de escudo representa un éxito y la de la otra cara un fracaso, entonces el conjunto $A = \{(0, 0, 0, 0, 0)\}$ constituye el suceso que consiste en que el escudo no aparece en ninguno de los 5 lanzamientos. La probabilidad del suceso A^c – el escudo aparece al menos una vez se calcula

$$P(A^c) = 1 - P(A) = 1 - 1/32 = 31/32 \approx 0,969.$$

■

La descripción del conjunto Ω de resultados de un experimento, y la de los resultados correspondientes a un suceso de interés A , puede ser una tarea engorrosa, como sucede cuando se trata de variaciones. Pero cuando Ω es finito, usualmente en el cálculo de probabilidades no se requiere tal descripción, sino que es suficiente conocer el número de elementos de los conjuntos Ω y A .

Ejemplo 3.7 *Problema de los cumpleaños.* Un grupo tiene r ($r \leq 365$) estudiantes. Hallemos una fórmula para la probabilidad p_r de que coincida el día de cumpleaños de al menos dos estudiantes del grupo. Se asumirá que cada nacimiento ocurre en cada día del año con igual probabilidad y se ignorará el 29 de febrero.

Sea A el suceso que consiste en la coincidencia del cumpleaños de al menos dos miembros del grupo. El suceso A^c significa que no coinciden ninguno de los cumpleaños en el grupo. Para calcular $P(A)$ usaremos la fórmula $P(A) = 1 - P(A^c)$.

No coinciden ninguno de los cumpleaños en el grupo cuando: un estudiante dado cumple años un día de los 365; un segundo estudiante cumple un día diferente de entre los 364 días restantes; un tercer estudiante cumple un día diferente de los anteriores, de entre los 363 restantes; y así sucesivamente. Luego el número de elementos de A^c se calcula por la fórmula,

$$|A^c| = 365 \times 364 \times \cdots \times (365 - (r - 1)) = V_{365}^r.$$

El conjunto Ω de resultados igualmente probables contiene 365^r elementos. Entonces

$$P(A^c) = \frac{365 \times 364 \times \cdots \times (365 - (r - 1))}{365^r}.$$

Por consiguiente,

$$\begin{aligned} p_r &= P(A) = 1 - P(A^c) = 1 - \left(\frac{365 - 1}{365} \right) \left(\frac{365 - 2}{365} \right) \cdots \left(\frac{365 - (r - 1)}{365} \right) \\ &= 1 - (1 - 1/365)(1 - 2/365) \cdots (1 - (r - 1)/365). \end{aligned}$$

Para $r = 23$ y $r = 30$ se halla que $p_{23} = 0,507$ y $p_{30} = 0,706$, respectivamente. Un año tiene 365 días, por lo que, a primera vista, con grupos de solo 23 y 30 estudiantes, podría parecer que para estos casos las probabilidades correspondientes no deberían ser tan altas. Pero note que existen

$$\binom{r}{2} = \frac{r(r - 1)}{2}$$

pares de estudiantes para los cuales se comparan sus cumpleaños. Con $r = 30$ son 435 pares. ■

Ejemplo 3.8 Un lote de N artículos contiene M artículos defectuosos, y por lo tanto, $N - M$ no defectuosos. Se selecciona al azar, sin reposición, una

muestra de n artículos del total de N . Hallemos la probabilidad $P_n(k; N, M)$ de que entre los n artículos elegidos haya exactamente k defectuosos.

Naturalmente que $0 \leq k \leq \min(n, M)$. Si se consideran los n artículos seleccionados sin importar el orden en que estos se eligen, entonces el total de muestras de tamaño n igualmente probables es $\binom{N}{n}$. Por otro lado, k artículos defectuosos se pueden seleccionar entre los M defectuosos de $\binom{M}{k}$ formas, y el número de formas en que los restantes $n - k$ artículos se pueden elegir entre los no defectuosos, hasta completar la muestra de tamaño n , es $\binom{N-M}{n-k}$. Entonces el número de muestras, correspondientes al suceso que consiste en que k de los n artículos seleccionados son defectuosos, es igual a

$$\binom{M}{k} \binom{N-M}{n-k}.$$

De donde,

$$P_n(k; N, M) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Esta probabilidad también puede calcularse considerando los diferentes órdenes en que los integrantes de una muestra de n artículos pueden ser elegidos. En este caso, V_N^n es el total de muestras posibles igualmente probables en la selección de n artículos del total de N . Por otra parte, para calcular el número de estas muestras que contienen k artículos defectuosos, observemos lo siguiente: el número de muestras de tamaño n que se pueden formar con k artículos es $\binom{n}{k}$; a cada una de estas muestras corresponden V_M^k formas posibles de seleccionar k artículos defectuosos de entre el total de M defectuosos y V_{N-M}^{n-k} formas de elegir $n - k$ no defectuosos de entre los $N - M$ no defectuosos. Luego,

$$P_n(k; N, M) = \frac{\binom{n}{k} V_M^k V_{N-M}^{n-k}}{V_N^n},$$

resultado, que como es obvio y puede comprobarse, coincide con el obtenido anteriormente. ■

Ejemplo 3.9 Supongamos que en las condiciones del ejemplo anterior, se cambia la selección sin reposición por una selección con reposición. Es decir, después de la elección de un artículo, este se devuelve al lote antes de la elección del próximo. Ahora, con esta nueva condición, la probabilidad $P_n(k; N, M)$ se calcula por la fórmula

$$P_n(k; N, M) = \frac{\binom{n}{k} M^k (N-M)^{n-k}}{N^n} = \binom{n}{k} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}.$$



Observe que en los dos últimos ejemplos, similar a varios ejemplos anteriores, el cálculo de la probabilidad $P_n(k; N, M)$ se realizó sin describir el conjunto de resultados del experimento y tampoco el conjunto de resultados asociados a la ocurrencia del suceso de interés.

3.4. Problemas y ejercicios

1. Un matrimonio tiene cuatro hijos. Se asume que la probabilidad de que un hijo sea varón es igual a la probabilidad de que este sea mujer. ¿Qué es más probable, que dos hijos sean del mismo sexo, o que tres sean de un sexo y uno del otro sexo? Calcule las probabilidades de ambos casos. (A primera vista podría parecer que la probabilidad de que dos hijos sean varones y dos sean mujeres es $1/2$).
2. Una moneda equilibrada se lanza n veces al azar. Para $n = 3$ y $n = 4$, calcule las probabilidades de que el escudo aparezca un número impar y un número par de veces.
3. *Dos jugadores realizan n jugadas. El juego es justo, es decir, los jugadores tienen igual probabilidad de ganar en cada jugada; además, no se hacen tablas.
 - a) ¿Cuál es la probabilidad de que un jugador dado gane k jugadas, $0 \leq k \leq n$?
 - b) ¿Cuál es la probabilidad de que al finalizar el juego un jugador dado aventaje al otro en k jugadas ganadas, $0 \leq k \leq n$?
 - c) Implemente en lenguaje Python un programa de simulación, que permita calcular aproximadamente la probabilidad del inciso anterior para $n = 20$ y $k = 8$. Compare el resultado con el valor exacto.
4. Un profesor entrega a un grupo de estudiantes 10 problemas de los cuales se seleccionarán de manera aleatoria 5 de ellos, que consistirán en el examen. Si un estudiante sabe resolver 7 problemas, ¿cuál es la probabilidad de que el estudiante resuelva correctamente (con ello aprueba) al menos 3 problemas? **R:** $11/12$.

5. Considere el *Problema de los cumpleaños* del Ejemplo 3.7. Elabore un programa en lenguaje Python para calcular la probabilidad p_r de que coincida el día de cumpleaños de al menos dos estudiantes del grupo, para $2 \leq r \leq 40$.
6. Considere el *Problema de los cumpleaños* del Ejemplo 3.7. ¿Qué cantidad mínima de estudiantes tendría que tener un grupo, al cual usted pertenece, para que la probabilidad de que algún otro estudiante tenga el mismo día del cumpleaños suyo sea mayor que 0,5? **R:** 253.
7. Se realiza una serie de n experimentos, en cada uno de los cuales únicamente ocurre uno y solo uno de tres sucesos A_1 , A_2 y A_3 que son igualmente probables. Obtenga una fórmula para la probabilidad de que A_1 , A_2 y A_3 ocurran en k_1 , k_2 y k_3 experimentos, respectivamente.
8. * Un lote de N artículos contiene M defectuosos. Se seleccionan sucesivamente, y sin reposición, n artículos al azar. Calcule la probabilidad de que el i -ésimo seleccionado sea defectuoso, $1 \leq i \leq n$. **R:** M/N .

SOBRE DECISIÓN ESTADÍSTICA

Como se señaló anteriormente, la teoría de las probabilidades y la estadística matemática tienen objetos de estudio diferentes; sin embargo, se interrelacionan en la teoría y en las aplicaciones.

Uno de los problemas de la estadística consiste en la estimación de alguna característica de una población de individuos u objetos, a partir de los datos de una muestra de estos. Por ejemplo, la estimación de la probabilidad de un suceso usando para ello la frecuencia relativa de su ocurrencia en una serie de experimentos. Otros ejemplos de problemas de estimación son los examinados en el Ejemplo 1.6.

En las siguientes dos secciones, mediante dos ejemplos, se examina parcialmente un segundo problema de la estadística, que consiste en tomar una decisión sobre alguna característica de una población a partir de los datos de una muestra suya.

4.1. Un problema de mercadeo

Los artículos A y B cumplen iguales funciones, y se conoce que durante una etapa inicial de ventas tenían similar preferencia del público. Pero transcurrido un tiempo aparece la sospecha, marcadamente, de que existe preferencia por el artículo A. Se tomará una muestra al azar de n consumidores de ambos productos y se les preguntará cuál es el de su preferencia.

Si la gran mayoría de los individuos de la muestra escogen el producto A, intuitivamente esto sugiere que también la población general prefiere A, es decir, que ya no existe la igualdad de preferencias. Como se trata de tomar una decisión a partir de una muestra, podría incurrirse en un error, puesto que por puro azar en la muestra pudieran haberse incluido más individuos que prefieren A.

Para tomar una decisión fundamentada, y valorar el posible error con esa decisión, se puede proceder de la siguiente manera. Se supone, como sucedía en la etapa inicial de ventas, que los artículos A y B tienen igual preferencia, y se establece una regla, basada en los datos de la muestra, por medio de la cual se tomará la decisión de rechazar, o no, esta hipótesis; se calcula además la probabilidad, que denotaremos por p , de incurrir en un error con esta decisión, es decir, la probabilidad de decidir, según la regla, que los dos productos no tienen igual preferencia cuando en realidad esta es igual (o similar).

Si la probabilidad p es muy pequeña, esto significa que se puede rechazar la suposición de igualdad de preferencias. Pero si p no es pequeña, esta suposición no debería ser rechazada, debido a que es alta la probabilidad de una decisión errónea al hacerlo; en ese caso tampoco se debería concluir la igualdad de preferencias, sino solo que los datos de esa muestra no contradicen esta igualdad, ya que con otra muestra o una mayor cantidad de datos, la suposición podría ser rechazada.

Señalemos que desde el punto de vista práctico, si no se rechaza la igualdad de preferencias, esta se acepta.

Por otra parte, si se aceptara la hipótesis de la igualdad de preferencias, debería estimarse la probabilidad de incurrir en otro tipo de error, que consiste en decidir que existe igualdad de preferencias cuando en realidad es mayor la preferencia por A. En este texto no analizaremos las cuestiones relacionadas con la probabilidad de este tipo de error.

Supongamos que $n = 15$. Como se sospecha predilección por A, entonces una posible **regla de decisión estadística**, que parece adecuada para rechazar la suposición, presumiblemente falsa, de que las preferencias por los artículos son iguales, es la siguiente: esta hipótesis será rechazada, si la cantidad de consumidores que optan por A es superior a 11.

Calculemos ahora la probabilidad p de incurrir en un error con esta regla. El número de resultados con los cuales se rechaza la hipótesis de la igualdad

4.2. Un ejemplo sobre control de calidad

de preferencias, según la regla asumida, es el siguiente:

$$\binom{15}{12} + \binom{15}{13} + \binom{15}{14} + \binom{15}{15} = 455 + 105 + 15 + 1 = 576.$$

El total de resultados posibles en la muestra es igual a 2^{15} y, bajo la suposición de la igualdad de preferencias, estos resultados son equiprobables. Por consiguiente,

$$p = 576/2^{15} = 576/32768 \approx 0,017.$$

Este resultado se interpreta de la siguiente manera. En las condiciones del problema analizado, si, por ejemplo, se utilizara la regla de decisión propuesta en 1000 ocasiones, entonces en aproximadamente 17 de estas ocasiones se incurriría en error al rechazar la igualdad de preferencias.

En el ejemplo, el valor de p obtenido podría considerarse pequeño o no en dependencia de la trascendencia que tendría incurrir en un error al tomar la decisión de rechazar la hipótesis según la regla asumida. En general, los valores de n y p , y las reglas de decisión, dependen de las características del problema práctico específico que se esté estudiando. En el mismo ejemplo, con $n = 15$, si la regla de decisión consiste en rechazar la hipótesis cuando la cantidad de consumidores que optan por A es mayor que 13, entonces se obtiene una probabilidad p menor. También aumentando el tamaño n de la muestra se pueden establecer reglas de decisión con valores menores de p .

4.2. Un ejemplo sobre control de calidad

Consideremos un lote de N artículos que contiene un número M , desconocido, de artículos defectuosos. Con el propósito de evaluar la calidad del lote se selecciona al azar una muestra de n artículos. Si $n \leq N - M$, para la probabilidad $P_n(0, N, M)$, de que la muestra no contenga artículos defectuosos, se halla que (vea el Ejemplo 3.8)

$$\begin{aligned} P_n(0, N, M) &= \frac{\binom{N-M}{n}}{\binom{N}{n}} = \frac{(N-M)(N-M-1) \cdots (N-M-n+1)}{N(N-1) \cdots (N-n+1)} \\ &= \left(1 - \frac{M}{N}\right) \left(1 - \frac{M}{N-1}\right) \cdots \left(1 - \frac{M}{N-n+1}\right). \end{aligned}$$

Consideremos ahora una regla de decisión según la cual el lote se rechaza o no; se rechaza si la muestra contiene al menos un artículo defectuoso, y no se rechaza si esta no contiene artículos defectuosos. Entonces $P_n(0, N, M)$ es

la probabilidad de no rechazar el lote cuando contiene M artículos defectuosos. Supongamos que el lote se considera “malo” si contiene al menos M_0 artículos defectuosos, o sea si $M \geq M_0$, y “bueno” cuando $M < M_0$. Estime-mos la probabilidad de cometer el error de aceptar un lote “malo” según esta regla.

Observe que $P_n(0, N, M)$ es una función decreciente de M y entonces

$$P_n(0, N, M) < P_n(0, N, M_0 - 1), \text{ si } M \geq M_0.$$

Luego $P_n(0, N, M_0 - 1)$ es una cota superior de la probabilidad de aceptar el lote siendo “malo”, para todo valor de M tal que $M \geq M_0$.

Sean, por ejemplo, $N = 100$, $n = 35$ y $M_0 = 7$. Con este valor de M_0 , un lote se considera “malo” si contiene al menos 7 artículos defectuosos. Para estos valores, la probabilidad de incurrir en el error de no rechazar un lote “malo”, usando la regla de decisión asumida, es menor que

$$P_{35}(0, 100, 6) = 0,06929,$$

para todo $M \geq M_0$. Para valores de M_0 mayores que 7 la cota que se obtiene para estimar la probabilidad de error es menor.

Si en la muestra se encuentra al menos un artículo defectuoso, el lote se rechaza, y en este caso puede incurrirse en el error de rechazar el lote aun cuando este es “bueno”, esto es, cuando $M < M_0$. Como se mencionó en la sección anterior, en este texto no se analizará la probabilidad de este tipo de error

4.3. Problemas y ejercicios

1. Una persona dice tener percepción extrasensorial. Para verificar si esto es cierto, se lanza 10 veces al azar una moneda equilibrada y se le pide a la persona que, antes de cada lanzamiento, prediga el resultado que se obtendrá. El número de aciertos fue 9. Halle la probabilidad de que respondiendo al azar se pueda lograr un resultado al menos tan bueno como el obtenido. Comente sobre la probabilidad hallada.
2. Se sospecha que una moneda no está equilibrada. Para verificar esto, se realiza un experimento que consiste en lanzarla n veces al azar. Proponga una regla de decisión que, sobre la base de los resultados del experimento, permita rechazar la hipótesis de que la moneda está equilibrada, en los siguientes casos:

a) $n = 15$ y i) $p \leq 0,05$, ii) $p \leq 0,01$

b) $n = 10$ y $p \leq 0,05$,

donde p denota la probabilidad de rechazar erróneamente la hipótesis.

3. Considere un lote que contiene 100 artículos. El lote se considera “malo” si al menos 10 de sus artículos son defectuosos. Suponga que se establece una regla de decisión, según la cual el lote no se rechaza si en una muestra de artículos suyos no hay defectuosos. Estime el tamaño mínimo de la muestra que garantiza una probabilidad menor que 0,05 de no rechazar el lote siendo “malo”. **R:** 28.

DEFINICIÓN DE PROBABILIDAD Y PROPIEDADES

En este capítulo se comienza la construcción axiomática de la teoría de las probabilidades y se presentan las propiedades más inmediatas que se deducen de los axiomas.

5.1. Sucesos y sigma-álgebra

Recordemos que el conjunto de resultados de un experimento aleatorio se identifica con un conjunto denotado por Ω llamado espacio muestral, y que los sucesos que podrían ser observados en el experimento se identifican con subconjuntos de Ω .

Veamos el lenguaje utilizado con los sucesos y las principales operaciones entre estos. Las operaciones con sucesos son operaciones con conjuntos.

Ya se conoce que al conjunto $A^c = \Omega \setminus A$ se le llama **suceso contrario** del suceso A y que representa la no ocurrencia de A .

Sean $A, B \subset \Omega$. Si A y B son conjuntos tales que $B \subset A$, entonces se dice que la ocurrencia de B **implica la ocurrencia** de A .

Sean $A, B \subset \Omega$. El conjunto $A \cap B$ representa la ocurrencia de los sucesos A y B , y el conjunto $A \cup B$ la ocurrencia de al menos uno de los sucesos A

o B .

Si los conjuntos A y B son disjuntos, es decir, $A \cap B = \emptyset$, se dice que los sucesos A y B son **excluyentes**. En este caso, al realizar el experimento es imposible la ocurrencia de ambos sucesos. Más general, si los subconjuntos de Ω , A_1, A_2, \dots , cumplen que $A_i \cap A_j = \emptyset$, para todos $i \neq j, i, j \geq 1$, entonces se dice que estos sucesos son **mutuamente excluyentes**.

Note que dos sucesos contrarios cualesquiera, A y A^c , son excluyentes y, por otra parte, $A \cup A^c = \Omega$, lo cual significa que al realizar el experimento siempre ocurre al menos uno (pero solo uno) de los sucesos A o A^c .

Ejemplo 5.1 Sea $\Omega = \{1, 2, 3, 4, 5, 6\}$ el espacio muestral correspondiente al experimento que consiste en el lanzamiento al azar de un dado. El conjunto $A = \{1, 3, 5\}$ representa el suceso que consiste en que el resultado del experimento es un número impar, luego A^c significa que el resultado es un número par. El conjunto $B = \{5, 6\}$ representa que el resultado del experimento es un número mayor que cuatro.

Si al lanzar el dado apareció el número 5, entonces se dice que ocurrieron los sucesos A y B , o sea, el suceso $A \cap B$; si apareció uno de los números del conjunto $\{1, 3, 5, 6\}$, se dice que ocurrió al menos uno de los sucesos A o B , es decir, ocurrió $A \cup B$. La ocurrencia de $C = \{1, 5\}$ implica la ocurrencia de A , pues $C \subset A$. ■

Ejemplo 5.2 Cuando observamos el tiempo de vida en horas de una unidad, equipo, u organismo vivo, podría considerarse como espacio muestral el intervalo $\Omega = [0, +\infty)$. Sucesos de interés pueden ser los representados por los conjuntos

$$A = [0, 500), \quad B = (1000, +\infty), \quad C = [500, 1000],$$

que representan, respectivamente, que el tiempo de vida es menor que 500 horas, que es mayor que 1000 horas, y que la vida se extingue en el rango de tiempo comprendido entre las 500 y 1000 horas. Los sucesos A , B y C son excluyentes. Por otro lado, como $B \subset A^c$, entonces B implica A^c , o sea, si el tiempo de vida es mayor que 1000 horas, esto implica que es mayor o igual que 500 horas. ■

La probabilidad será definida sobre una colección \mathcal{F} de subconjuntos de Ω , cerrada por las operaciones de conjuntos finitas e infinitas numerables, la cual se define a continuación.

Definición 1.

Sea Ω un conjunto no vacío arbitrario. Una colección \mathcal{F} de subconjuntos de Ω se llama **σ -álgebra** en Ω si cumple las siguientes propiedades:

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

El par (Ω, \mathcal{F}) se llama **espacio medible**.

De esta definición se deduce inmediatamente que si \mathcal{F} es una σ -álgebra, entonces $\emptyset \in \mathcal{F}$, y que toda σ -álgebra es cerrada por las uniones e intersecciones finitas e infinitas numerables de elementos de esta; además, que si $A, B \in \mathcal{F}$ entonces $A \setminus B \in \mathcal{F}$.

Ejemplo 5.3 Dado un espacio muestral Ω , los siguientes tres conjuntos constituyen σ -álgebras en Ω :

1. $\mathcal{F}_0 = \{\emptyset, \Omega\}$,
2. $\mathcal{F}_1 = \{\emptyset, \Omega, A, A^c\}$, $A \subset \Omega$,
3. $\mathcal{F}_2 = \mathcal{P}(\Omega)$, donde $\mathcal{P}(\Omega)$ es el conjunto potencia de Ω .

A \mathcal{F}_0 se le llama *σ -álgebra trivial* y a \mathcal{F}_2 *σ -álgebra total*. ■

Ejemplo 5.4 Sea $\Omega = \{1, 2, 3, 4\}$. La colección

$$\mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2\}, \{3, 4\}, \{1, 2\}, \{1, 3, 4\}, \{2, 3, 4\}\},$$

es una σ -álgebra en Ω . ■

Ejercicio 1.

Considere una colección arbitraria de σ -álgebras en Ω . Pruebe que la intersección de estas σ -álgebras es también una σ -álgebra en Ω . ¿Es la unión de σ -álgebras una σ -álgebra?

Ejercicio 2.

Sea el espacio medible (Ω, \mathcal{F}) , y $A \subset \Omega$. Pruebe que

$$\mathcal{F}_A = \{C : C = B \cap A, B \in \mathcal{F}\}$$

es una σ -álgebra en A .

Definición 2.

Dada una colección arbitraria \mathcal{C} de subconjuntos de un conjunto no vacío E , se llama **σ -álgebra generada** por \mathcal{C} , o *mínima σ -álgebra que contiene a \mathcal{C}* , y se denota por $\sigma(\mathcal{C})$, a la intersección de todas las σ -álgebras que contienen a \mathcal{C} . El conjunto \mathcal{C} se llama **sistema generador** de $\sigma(\mathcal{C})$.

Pueden existir diferentes sistemas generadores de una misma σ -álgebra.

Ejemplo 5.5 Si $A \subset E$ y $\mathcal{C} = \{A\}$, entonces $\sigma(\mathcal{C}) = \sigma(\{A^c\}) = \{E, \emptyset, A, A^c\}$.

Si $E = \{1, 2, 3, 4\}$, $\mathcal{C}_1 = \{\{1\}, \{3, 4\}\}$ y $\mathcal{C}_2 = \{\{2\}, \{1, 2\}\}$, entonces

$$\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2) = \{E, \emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4\}, \{2, 3, 4\}, \{1, 3, 4\}\}.$$

■

5.2. Definición de probabilidad

Definición 3 (Axiomas de Kolmogorov).

Una función real P definida sobre la σ -álgebra \mathcal{F} en Ω se denomina **probabilidad** (o medida de probabilidad) si cumple las siguientes propiedades o axiomas:

1. $P(A) \geq 0$ para todo $A \in \mathcal{F}$
2. $P(\Omega) = 1$
3. Si $A_1, A_2, \dots \in \mathcal{F}$ y $A_i \cap A_j = \emptyset$, $i \neq j$, $i, j \geq 1$, entonces

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

El número $P(A)$ se llama probabilidad de A .

A veces, cuando se hace referencia a una probabilidad P definida sobre una σ -álgebra \mathcal{F} en Ω , se dice que esta probabilidad está definida sobre el espacio medible (Ω, \mathcal{F}) .

La probabilidad P mide cada elemento de la σ -álgebra \mathcal{F} en la cual está definida, en el sentido de que le asigna un valor numérico. Todo elemento de \mathcal{F} se dice que es un **conjunto medible**.

Hasta ahora nos hemos referido a los sucesos como subconjuntos de un espacio muestral Ω , sin precisar que en realidad no cualquier subconjunto de Ω es un suceso. En la teoría de las probabilidades solo son **sucesos** o **eventos** los elementos de la σ -álgebra \mathcal{F} sobre la cual se define la probabilidad, es decir, los sucesos son los conjuntos medibles.

Por otro lado, dado un espacio muestral Ω infinito no numerable, no sobre cualquier σ -álgebra en Ω será posible definir una determinada probabilidad P . La σ -álgebra a considerar deberá contener los sucesos con significado teórico y práctico, y además tendrá que ser posible definir sobre ella la probabilidad P . Por ejemplo, la probabilidad geométrica no se puede definir sobre la colección de todos los subconjuntos del intervalo $[0, 1]$, puesto que, como se verá más adelante, no todo subconjunto de este intervalo tiene

longitud.

El axioma 3 de la definición de probabilidad se denomina **σ -aditividad** o **aditividad numerable** de la probabilidad; también se dice que la probabilidad P es σ -aditiva. Una interpretación de este axioma es la siguiente. Consideremos un subconjunto Ω de \mathbb{R}^n que tiene volumen igual a 1, y una σ -álgebra \mathcal{F} de subconjuntos de Ω para los cuales está definido el volumen. Interpretemos el volumen de cada elemento de \mathcal{F} como su probabilidad. En este caso el axioma de σ -aditividad significa, que dada una partición infinita numerable de un elemento de \mathcal{F} en elementos también de \mathcal{F} , el volumen del conjunto particionado es igual a la suma infinita de los volúmenes de los conjuntos que constituyen la partición.

Pero recordemos que existen subconjuntos de \mathbb{R}^n que no tienen volumen definido en el sentido de la integral de Riemann, y, sin embargo, cada elemento de una partición infinita numerable de los mismos tiene volumen finito.

En la Sección 10.2, se examina una σ -álgebra de subconjuntos de \mathbb{R}^n , llamada σ -álgebra de Borel, sobre la cual se define una *medida* denominada medida de Lebesgue, denotada por λ , que coincide con la noción de volumen y es σ -aditiva. Pero no todo subconjunto de \mathbb{R}^n tiene volumen según λ ; esto se muestra en la Sección 8.1, donde se presenta un subconjunto del intervalo $[0, 1]$ que no es medible según λ , es decir, que no tiene longitud. El examen cabal de estas cuestiones corresponde a la *teoría de la medida*.

Señalemos que en una definición rigurosa de la probabilidad geométrica, el volumen debe considerarse según la medida λ definida sobre la σ -álgebra de Borel. Entonces la probabilidad geométrica cumple con los axiomas de probabilidad.

En el axioma 3, la unión de conjuntos no puede ser infinita no numerable; esto se ilustra con el siguiente ejemplo. Sea $\Omega = [0, 1]$ y consideremos la probabilidad geométrica. En este caso, con la extensión del axioma 3 a la unión no numerable se obtendría el siguiente absurdo:

$$1 = \lambda([0, 1]) = \sum_{x \in [0, 1]} \lambda(\{x\}) = 0.$$

Por otro lado, como se verá posteriormente, tanto por razones teóricas como prácticas debido a operaciones con límites, el axioma 3 debe establecer la σ -aditividad y no solo la aditividad finita.

Destaquemos que en la construcción del *modelo probabilístico* teórico de un experimento se seleccionan, el espacio muestral Ω , la σ -álgebra \mathcal{F} y la probabilidad P que lo modelan, es decir, se selecciona un espacio medible (Ω, \mathcal{F})

que contiene la información de interés sobre el experimento y en el cual es posible definir la probabilidad P . Un modelo probabilístico se representa por la terna (Ω, \mathcal{F}, P) , denominada **espacio de probabilidad**. La teoría de las probabilidades estudia modelos probabilísticos.

5.3. Propiedades básicas de la probabilidad

En el siguiente teorema se formulan propiedades básicas de la probabilidad que se deducen directamente de los axiomas.

Teorema 1.

Sean el espacio de probabilidad (Ω, \mathcal{F}, P) y $A, B, A_i \in \mathcal{F}, i \geq 1$. Entonces

1. $P(\emptyset) = 0$
2. $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$, si $A_i \cap A_j = \emptyset, i \neq j, 1 \leq i, j \leq n$
3. $P(A^c) = 1 - P(A)$
4. Si $B \subset A$ entonces $P(A \setminus B) = P(A) - P(B)$ y $P(B) \leq P(A)$
5. $0 \leq P(A) \leq 1$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$
8. $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

Demostración:

Probemos las propiedades 1 y 8. La demostración del resto de las propiedades se propone como ejercicio.

Por los axiomas 2 y 3 se tiene que

$$1 = P(\Omega) = P(\Omega \cup \emptyset \cup \dots \cup \emptyset) = 1 + P(\emptyset) + P(\emptyset) + \dots,$$

de donde $P(\emptyset) = 0$.

Probemos ahora la propiedad 8. Tomemos (vea la Figura 5.1)

$$B_1 = A_1, \quad B_i = A_i \setminus \bigcup_{k=1}^{i-1} A_k, \quad i \geq 2.$$

Se tiene que $B_i \cap B_j = \emptyset$, para todos $i \neq j$, $i, j \geq 1$, y

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i, \quad B_i \subset A_i, \quad i \geq 1.$$

Por la aditividad numerable de la probabilidad y la propiedad 4 del teorema, obtenemos,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

■

Las propiedad 2 del Teorema 1 se denomina *aditividad finita* de la probabilidad, y las propiedades 7 y 8, *sub-aditividad finita* y *sub-aditividad numerable* de la probabilidad, respectivamente.

Anteriormente se señaló que la probabilidad geométrica cumple con la aditividad numerable. Note ahora que la definición clásica de probabilidad y la frecuencia relativa de ocurrencia de un suceso, cumplen con los axiomas 1 y 2, y con la propiedad de aditividad finita.

Es importante puntualizar, que existen conjuntos diferentes de los sucesos Ω y \emptyset con probabilidades iguales a 1 o a 0. Por ejemplo, si se consideran,

$$\Omega = [0, 1] \times [0, 1] \quad \text{y} \quad A = \{1/2\} \times [0, 1],$$

y la probabilidad geométrica λ , entonces $\lambda(A) = 0$ y $\lambda(\Omega \setminus A) = 1$.

Ejercicio 3.

Demuestre las propiedades 2-7 del Teorema 1.

Ejemplo 5.6 Una moneda se lanza al azar dos veces. Como anteriormente, la aparición del escudo se representará por el número 1 y la de la otra cara

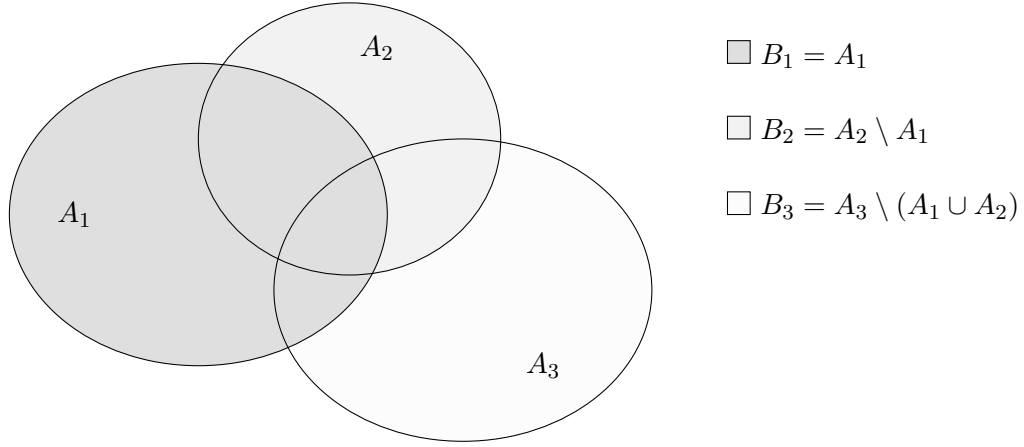


Figura 5.1: Diagrama para la demostración en el Teorema 1.

por 0, y consideraremos pares ordenados con componentes 0 y 1, donde la primera componente de cada par se asocia al resultado correspondiente al primer lanzamiento, y su segunda componente al resultado correspondiente al segundo lanzamiento.

Un espacio muestral relacionado con los dos lanzamientos es el conjunto

$$\Omega = \{\omega = (x_1, x_2) : x_1, x_2 \in \{0, 1\}\}.$$

Sean

$$\begin{aligned} A_1 &= \{(1, x_2) : x_2 \in \{0, 1\}\} \\ A_2 &= \{(x_1, 1) : x_1 \in \{0, 1\}\}. \end{aligned}$$

Los sucesos A_1 y A_2 representan la aparición de escudo en el primero y segundo lanzamiento de la moneda, respectivamente. Calculemos la probabilidad del suceso B , que consiste en que el escudo aparece en al menos uno de los dos lanzamientos, o sea $B = A_1 \cup A_2$. Si la moneda es equilibrada, se puede asumir que los cuatro puntos muestrales son igualmente posibles. Usando la propiedad 6 del Teorema 1 se halla que

$$\begin{aligned} P(B) &= P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \\ &= 1/2 + 1/2 - 1/4 = 3/4. \end{aligned}$$

Claro que $P(B)$ puede calcularse directamente sin usar la propiedad 6.

■

En el ejemplo anterior, el cálculo de $P(B)$ cuando la moneda no está equilibrada se relaciona con el concepto de independencia de sucesos, el cual será examinado posteriormente.

5.4. Modelo discreto

Haciendo una identificación entre el conjunto puntual $\{\omega\}$, constituido por un único punto muestral, y este punto muestral ω , denotaremos los conjuntos puntuales simplemente por ω en lugar de $\{\omega\}$. Con esta notación, en lugar de $P(\{\omega\})$, en lo adelante se escribirá $P(\omega)$.

Sean el espacio muestral infinito numerable $\Omega = \{\omega_1, \omega_2, \dots\}$ y $\mathcal{F} = \mathcal{P}(\Omega)$. Una probabilidad P sobre \mathcal{F} puede definirse de la siguiente manera:

$$P(\omega_i) = p_i \geq 0, \quad i = 1, 2, \dots, \quad \sum_{i=1}^{\infty} p_i = 1,$$

y

$$P(A) = \sum_{i: \omega_i \in A} p_i, \quad \text{para todo } A \subset \Omega.$$

Además, $P(\emptyset) = 0$. En el caso en que Ω es finito la definición es análoga. Es fácil verificar que esta definición cumple con los axiomas de Kolmogorov.

Cuando el espacio muestral Ω es finito o infinito numerable diremos que este es discreto, y el espacio de probabilidad (Ω, \mathcal{F}, P) , donde $\mathcal{F} = \mathcal{P}(\Omega)$, y P se define como anteriormente, se denomina **modelo discreto**. En este modelo todos los subconjuntos de Ω son medibles (sucesos).

La definición clásica de probabilidad es un caso particular de modelo discreto en el cual $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$,

$$P(\omega_i) = 1/N, \quad i = 1, 2, \dots, N$$

$$P(A) = \sum_{i: \omega_i \in A} P(\omega_i) = \frac{|A|}{|\Omega|}, \quad \text{si } A \subset \Omega.$$

La frase “seleccionar un punto al azar” de un conjunto finito, significará que se asume el modelo discreto correspondiente a la definición clásica de probabilidad.

Ejercicio 4.

Pruebe que *toda* probabilidad sobre (Ω, \mathcal{F}) , donde Ω es discreto y $\mathcal{F} = \mathcal{P}(\Omega)$, está definida tal como se definió para el modelo discreto.

Ejemplo 5.7 Se lanza sucesivamente al azar una moneda equilibrada hasta la aparición de escudo (éxito) por primera vez. En este caso puede considerarse el espacio muestral $\Omega = \{1, 2, 3, \dots, \infty\}$, donde el punto muestral i , corresponde a la ocurrencia de éxito por primera vez en el i -ésimo lanzamiento, $i = 1, 2, \dots$, y el símbolo ∞ corresponde a que el éxito no ocurre nunca.

Puesto que la moneda es equilibrada, si esta se lanza solo un número finito de veces n , entonces la probabilidad de que el éxito ocurra por primera vez en el lanzamiento con número de orden i se calcula,

$$P(\{(0, \dots, 0, 1, x_{i+1}, x_{i+2}, \dots, x_n) : x_k \in \{0, 1\}, i+1 \leq k \leq n\}) = \frac{2^{n-i}}{2^n} = \frac{1}{2^i}.$$

Ahora podemos asumir, intuitivamente, que en una sucesión infinita de lanzamientos la probabilidad de que el éxito ocurra por primera vez en el i -ésimo lanzamiento es igual a $P(i) = 1/2^i$, $i \geq 1$. Entonces la probabilidad de que el éxito ocurra alguna vez en esta sucesión de lanzamientos es igual a la suma

$$\sum_{i=1}^{\infty} P(i) = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1.$$

De aquí se concluye que $P(\infty) = 0$, es decir, la probabilidad de que nunca ocurra éxito (siempre ocurra fracaso) es igual a 0. Desde luego que la probabilidad de que ocurra al menos un fracaso en la sucesión de lanzamientos también es igual a 1.

El subconjunto de Ω , $A = \{1, 3, 5, \dots\}$, se asocia con el suceso que consiste en que el éxito ocurre por primera vez en un lanzamiento con número de orden impar. La probabilidad de A se halla,

$$P(A) = \sum_{i=1}^{\infty} P(2i-1) = \sum_{i=1}^{\infty} \frac{1}{2^{2i-1}} = \frac{2}{3},$$

de donde, la probabilidad de que ocurra éxito por primera vez en un lanzamiento con número de orden par es igual a $1/3$, o sea, dos veces menor que la probabilidad de que el éxito ocurra en un lanzamiento impar. ■

Si el conjunto Ω contiene N puntos muestrales, el número de elementos del conjunto potencia $\mathcal{P}(\Omega)$ se calcula,

$$|\mathcal{P}(\Omega)| = \sum_{k=0}^N \binom{N}{k} = (1+1)^N = 2^N.$$

En el caso de una serie de n experimentos o pruebas, en cada uno de los cuales interesa solo la ocurrencia de éxito o fracaso, se tiene, $N = |\Omega| = 2^n$ y, por consiguiente, $|\mathcal{P}(\Omega)| = 2^{2^n}$. Si $n = 2$, se halla que $|\mathcal{P}(\Omega)| = 16$, y si $n = 3$, $|\mathcal{P}(\Omega)| = 256$.

Cuando en cada una de las n pruebas de la serie interesan no solo dos, sino tres sucesos excluyentes, y solo estos: éxito, fracaso, y ni éxito ni fracaso, entonces el espacio muestral contiene $N = |\Omega| = 3^n$ puntos muestrales y $|\mathcal{P}(\Omega)| = 2^{3^n}$. Si $n = 3$, se halla que $|\mathcal{P}(\Omega)| = 2^{3^3} = 2^{27} = 134217728$.

Cuando el espacio muestral Ω es infinito numerable, la σ -álgebra total $\mathcal{P}(\Omega)$ es no numerable.

En esta sección hemos visto que si el espacio muestral es discreto, la probabilidad de cualquier subconjunto suyo se define a partir de la probabilidad de los conjuntos puntuales; sin embargo, si el espacio muestral es infinito no numerable esto no es posible. Un ejemplo de esto es el caso de la probabilidad geométrica sobre el intervalo $\Omega = (0, 1)$, en el cual la probabilidad de cada conjunto puntual es igual a 0, aunque las probabilidades de numerosos subconjuntos de interés de Ω son diferentes de 0.

En la sección 8.2, se demuestra que en un espacio muestral infinito no numerable Ω , la probabilidad de los conjuntos puntuales es igual a 0, excepto quizás en los puntos de un subconjunto finito o infinito numerable de Ω .

5.5. Información contenida en un espacio medible

En esta sección se examinará parcialmente el sentido práctico de los espacios medibles.

Como se mencionó anteriormente, en la modelación de un experimento se seleccionan el espacio muestral Ω y la σ -álgebra \mathcal{F} en Ω que contienen la *información* de interés, o posible, sobre el resultado del experimento.

Una σ -álgebra $\mathcal{F}_1 = \{\emptyset, \Omega, A, A^c\}$, donde $A \subset \Omega$, contiene una *información mínima*, y modela experimentos a partir de cuya observación solo es posible decidir sobre la ocurrencia o no del suceso A .

La σ -álgebra total $\mathcal{F} = \mathcal{P}(\Omega)$, con espacios muestrales Ω finitos o infinitos numerables, contiene la *información máxima*, y modela experimentos a partir de cuyos resultados puede decidirse sobre la ocurrencia o no de cualquier suceso imaginable.

La σ -álgebra trivial $\mathcal{F}_0 = \{\Omega, \emptyset\}$ no contiene información; cuando se realiza el experimento, claro que siempre ocurre el suceso cierto Ω .

Consideremos en el Ejemplo 5.6, del lanzamiento dos veces de una moneda, las σ -álgebras

$$\mathcal{F}_1 = \{\Omega, \emptyset, A_1, A_1^c\}, \quad \mathcal{F}_2 = \{\Omega, \emptyset, A_2, A_2^c\}, \quad \mathcal{F}' = \{\Omega, \emptyset, B, B^c\}.$$

Las dos primeras σ -álgebras contienen información solo sobre la aparición o no de escudo en el primero y segundo lanzamiento, respectivamente, y la tercera contiene solo información sobre la aparición o no de escudo en al menos uno de los dos lanzamientos de la moneda. La σ -álgebra que contiene toda la información sobre los dos lanzamientos debe contener, en particular, los elementos de \mathcal{F}_1 y \mathcal{F}_2 , y es la σ -álgebra generada por la unión de esta dos σ -álgebras, $\sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$, la cual, en este caso, coincide con $\mathcal{P}(\Omega)$.

Como se vio en el Ejemplo 5.6, en dependencia del problema a resolver, la σ -álgebra a considerar no necesariamente tiene que ser la total. En ese ejemplo, si solo se requiere el cálculo de la probabilidad de éxito en el segundo lanzamiento, la σ -álgebra podría ser \mathcal{F}_2 .

Para un mismo experimento pueden construirse diferentes modelos, o sea diferentes espacios muestrales y diferentes σ -álgebras, de acuerdo a la información que pueda obtenerse del experimento y lo que se requiera calcular.

Ejercicio 5.

Sea el experimento que consiste en el lanzamiento de una moneda tres veces al azar. Construya una σ -álgebra que contenga toda la información hasta el segundo lanzamiento.

Cuando la sucesión de lanzamientos de la moneda es infinita, en general se considera el espacio muestral

$$\Omega = \{w = (x_1, x_2, \dots, x_n, \dots) : x_i \in \{0, 1\}, i = 1, 2, \dots\};$$

este conjunto es infinito no numerable. Sin embargo, en el Ejemplo 5.7 se vió que, si lo que interesa en el lanzamiento sucesivo de la moneda es el número de experimentos hasta que por primera vez aparece escudo, entonces puede considerarse el espacio muestral numerable $\Omega = \{1, 2, 3, \dots, \infty\}$.

5.6. Continuidad de la probabilidad

En lo adelante se considerará un espacio de probabilidad (Ω, \mathcal{F}, P) y que los sucesos pertenecen a la σ -álgebra \mathcal{F} .

Sea $\{A_n\}$ una sucesión de eventos pertenecientes a la σ -álgebra \mathcal{F} en Ω . El suceso que consiste en la *ocurrencia de infinitos sucesos* de esta sucesión se denota por $\{A_n \text{ i.f.}\}$, donde i.f significa **infinitamente frecuente**.

Note que $\omega \in \{A_n \text{ i.f.}\}$ si, y solo si, para todo $n, n \geq 1$, existe $k \geq n$ tal que $\omega \in A_k$. Entonces

$$\{A_n \text{ i.f.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Este conjunto se denota además por $\limsup_n A_n$, y también se llama **límite superior** de la sucesión de eventos $\{A_n\}$. Es decir,

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

El **límite inferior** de esta sucesión, denotado por $\liminf_n A_n$, se define

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

Ahora, $\omega \in \liminf_n A_n$ si, y solo si, existe $n, n \geq 1$, tal que para todo $k \geq n$ se tiene $\omega \in A_k$. En este caso el evento $\liminf_n A_n$ ocurre si, y solo si, ocurren todos los sucesos de la sucesión $\{A_n\}$, $n \geq 1$, con la excepción quizás de un número finito de estos. Observe que

$$\left(\limsup_n A_n \right)^c = \liminf_n A_n^c.$$

Cuando

$$\liminf_n A_n = \limsup_n A_n = A,$$

se dice que el límite de la sucesión existe y es igual a A , lo cual se denota, $\lim_n A_n = A$.

Note la analogía entre los conceptos de límite superior e inferior de una sucesión de eventos, y los conceptos de límite superior e inferior de una sucesión numérica.

Ejemplo 5.8 Sean dos sucesos B y C diferentes, y la sucesión $\{A_n\}$, tal que $A_{2n-1} = B$ y $A_{2n} = C$, $n \geq 1$. Entonces

$$\limsup_n A_n = B \cup C \quad \text{y} \quad \liminf_n A_n = B \cap C.$$

En este caso no existe el límite de la sucesión $\{A_n\}$. ■

Ejercicio 6.

Sea $\{A_n\}$ una sucesión de eventos. Pruebe que

$$\liminf_n A_n \subset \limsup_n A_n.$$

Una sucesión de eventos $\{A_n\}$, $n \geq 1$, se dice que es *creciente*, si $A_n \subset A_{n+1}$, para todo $n \geq 1$, y se dice que es *decreciente*, si $A_n \supset A_{n+1}$, para todo $n \geq 1$.

Ejercicio 7.

Demuestre que si una sucesión de eventos $\{A_n\}$ es creciente, entonces $\lim_n A_n$ existe y $\lim_n A_n = \bigcup_{k=1}^{\infty} A_k$, y que si la sucesión $\{A_n\}$ es decreciente, entonces $\lim_n A_n$ existe y $\lim_n A_n = \bigcap_{k=1}^{\infty} A_k$.

Teorema 2 (Continuidad de la Probabilidad).

Si $\{A_n\}$ es una sucesión creciente o decreciente de eventos, entonces

$$\lim_n P(A_n) = P(\lim_n A_n).$$

Demostración:

Supongamos que la sucesión $\{A_n\}$ es creciente, y definamos los sucesos $B_n, n \geq 1$, como (vea la Figura 5.2)

$$B_1 = A_1, \quad B_n = A_n \setminus A_{n-1}, \quad n > 1.$$

Los sucesos $B_n, n \geq 1$, son mutuamente excluyentes, $B_i \cap B_j = \emptyset, i \neq j, i, j \geq 1$. Además,

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i \quad \text{y} \quad A_n = \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i, \quad n \geq 1.$$

Luego,

$$\begin{aligned} P\left(\lim_n A_n\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_n \sum_{i=1}^n P(B_i) \\ &= \lim_n P\left(\bigcup_{i=1}^n B_i\right) = \lim_n P\left(\bigcup_{i=1}^n A_i\right) = \lim_n P(A_n). \end{aligned}$$

El caso de una sucesión $\{A_n\}$ decreciente se propone como ejercicio. ■

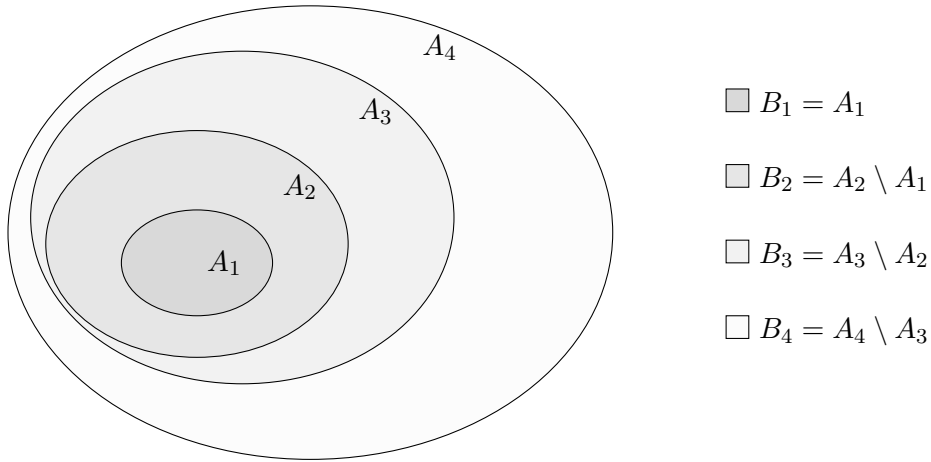


Figura 5.2: Diagrama para la demostración del Teorema 2.

Ejercicio 8.

Pruebe el Teorema 2 cuando la sucesión de eventos $\{A_n\}$ es decreciente.

Ejemplo 5.9 Consideremos el lanzamiento sucesivo de una moneda equilibrada. Sea A_n el suceso que consiste en que el escudo aparece en cada uno de los primeros n lanzamientos. Como

$$A_{n+1} \subset A_n \quad \text{y} \quad P(A_n) = 1/2^n, \quad n \geq 1,$$

entonces, por la continuidad de la probabilidad,

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_n P(A_n) = 0.$$

El suceso $\bigcap_{n=1}^{\infty} A_n$ significa que el escudo aparece en todos los lanzamientos, luego la probabilidad de que siempre aparezca escudo es igual a 0 y, por consiguiente, la probabilidad de que en algún lanzamiento aparezca la otra cara es igual a 1. En el Ejemplo 5.7 se llegó a esta conclusión por otra vía. ■

Ejercicio 9.

Sea $\{A_n\}$ una sucesión de eventos. Pruebe que:

1. $P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n)$.
2. Si $\lim_n A_n$ existe, entonces $\lim_n P(A_n) = P(\lim_n A_n)$.

La probabilidad del suceso $\{A_n \text{ i.f.}\}$ será denotada $P(A_n \text{ i.f.})$.

Ejemplo 5.10 Sea una sucesión de experimentos en cada uno de los cuales ocurre éxito con igual probabilidad p . Denotemos por A_n el suceso que consiste en la ocurrencia de éxito en el experimento con número de orden n , $n \geq 1$. Entonces, como $P(A_n) = p$, $n \geq 1$, de la primera parte del ejercicio anterior se deduce que $P(A_n \text{ i.f.}) \geq p$. En palabras, la probabilidad de que el éxito ocurra infinitamente frecuente es mayor o igual que p . ■

Ejercicio 10.

Sean A y B eventos y $\{A_n\}$ una sucesión de eventos. Demuestre la equivalencia entre los siguientes enunciados i y ii.

- i. a) $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \emptyset$;
 b) si $A_n \downarrow \emptyset$, entonces $\lim_n P(A_n) = 0$.

ii. P es σ -aditiva.

5.7. Lema de Borel-Cantelli. Parte I

Teorema 3 (Lema de Borel-Cantelli. Parte I.).

Sea $\{A_n\}$ una sucesión de eventos. Se cumple que:

$$\text{Si } \sum_{n=1}^{\infty} P(A_n) < \infty, \text{ entonces } P(A_n \text{ i.f.}) = 0.$$

Demostración:

Note que $\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$ es el límite de la sucesión decreciente $\bigcup_{k=n}^{\infty} A_k, n \geq 1$. Entonces, usando la continuidad de la probabilidad y la sub-aditividad numerable, se tiene que

$$\begin{aligned} P(A_n \text{ i.f.}) &= P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = P\left(\lim_n \bigcup_{k=n}^{\infty} A_k\right) \\ &= \lim_n P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_n \sum_{k=n}^{\infty} P(A_k) = 0. \end{aligned}$$

■

Ejemplo 5.11 Consideremos el Ejemplo 5.9 del lanzamiento sucesivo de una moneda equilibrada, donde A_n representa el suceso que consiste en que el escudo aparece en cada uno de los primeros n lanzamientos, $n \geq 1$. Como

$P(A_n) = 1/2^n$, $n \geq 1$, entonces $\sum_{k=1}^{\infty} P(A_k) < \infty$, de donde, por el lema de Borel-Cantelli, $P(A_n \text{ i.f.}) = 0$. Esto significa que, tal como se concluyó en los Ejemplos 5.7 y 5.9, la probabilidad de que siempre aparezca escudo es igual a 0. ■

Es evidente que si $\{A_n\}$ es una sucesión de eventos tales que $A_i \cap A_j = \emptyset$, para todos $i \neq j$, $i, j \geq 1$, entonces $P(A_n \text{ i.f.}) = 0$.

El recíproco del teorema anterior no se cumple en general, como muestra el siguiente ejemplo. Sean $\Omega = (0, 1)$, P la probabilidad geométrica y $A_n = (0, a_n)$, donde $a_n \downarrow 0$. Entonces $\limsup_n A_n = \bigcap_{n=1}^{\infty} A_n = \emptyset$, luego $P(A_n \text{ i.f.}) = P(\emptyset) = 0$. Sin embargo, cuando $a_n \geq 1/n$ se tiene

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

5.8. Problemas y ejercicios

1. Suponga que la sucesión de eventos $\{A_n\}$ es una partición numerable de Ω . Pruebe que la colección de todas las uniones de elementos de la partición, incluyendo además el conjunto \emptyset , es una σ -álgebra en Ω .
2. Verifique que si para los sucesos A y B se tiene, $P(A) = 3/4$ y $P(B) = 1/3$, entonces $1/12 \leq P(A \cap B) \leq 1/3$.
3. Pruebe que para eventos A_1, \dots, A_n arbitrarios se cumple la siguiente desigualdad (de Bonferroni)

$$P(A_1 \cap \dots \cap A_n) \geq \sum_{i=1}^n P(A_i) - (n-1) \quad n \geq 1.$$

4. *Un intervalo de longitud 1 se divide en n partes por $n-1$, $n \geq 2$, puntos de ese intervalo seleccionados al azar de manera independiente. ¿Cuál es la probabilidad de que todas las partes sean de longitud menor que $1/2$?

Sugerencia: Considere el suceso contrario del suceso cuya probabilidad se desea calcular. **R:** $1 - n(1/2)^{n-1}$, $n \geq 2$.

5. Sea $\{A_n\}$ una sucesión de eventos. Demuestre que $P(A_i) = 1$ para todo $i = 1, 2, \dots$ si, y solo si, $P(\bigcap_{i=1}^{\infty} A_i) = 1$.

6. En un grupo, el por ciento de estudiantes que aprobaron las asignaturas A , B y C es el siguiente: A : 50 %, B : 40 %, C : 30 %, A y B : 35 %, B y C : 20 %, A y C : 25 %, y 15 % aprobaron las tres asignaturas. Se selecciona un estudiante del grupo al azar ¿Cuál es la probabilidad de que haya aprobado al menos una de las tres asignaturas?

7. a) *Principio de inclusión-exclusión.* Sean A_1, A_2, \dots, A_n eventos. Pruebe que

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n P(A_k) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n). \quad (5.1)$$

- b) Los n estudiantes de un grupo colocan en una bolsa su libro de Álgebra y luego cada uno selecciona al azar un libro. Calcule la probabilidad p_n de que al menos un estudiante seleccione su libro. Halle $p = \lim_n p_n$.
R: $p = 1 - e^{-1}$.

8. Pruebe la propiedad de sub-aditividad numerable de la probabilidad usando la propiedad de continuidad de la probabilidad.
9. *Sea $\{A_n\}$ una sucesión de eventos. Demuestre la equivalencia de i, ii y iii.

- i. Si $A_n \downarrow$, entonces $\lim_n P(A_n) = P(\lim_n A_n)$.
- ii. Si $A_n \downarrow \emptyset$, entonces $\lim_n P(A_n) = 0$.
- iii. Si $A_n \uparrow$, entonces $\lim_n P(A_n) = P(\lim_n A_n)$.

PROBABILIDAD CONDICIONAL E INDEPENDENCIA DE SUCESOS

En este capítulo se estudia la probabilidad condicional, concepto que tiene gran importancia teórica y numerosas aplicaciones prácticas. La noción de independencia es crucial en la teoría de probabilidades y la estadística matemática; en el capítulo se inicia el estudio de este concepto a través de la independencia de sucesos.

6.1. Probabilidad condicional

Consideremos el lanzamiento al azar de un dado equilibrado. Supongamos que se tiene la información de que como resultado del lanzamiento del dado, el número que aparece (o apareció) es mayor que 4. ¿Cuál es la probabilidad de que este número sea impar?

Sean $\Omega = \{1, 2, 3, 4, 5, 6\}$ y los sucesos $A = \{1, 3, 5\}$ y $B = \{5, 6\}$. La probabilidad que se desea calcular es la probabilidad del suceso A , pero condicionada a que se posee la información de que el suceso B ocurre. Que A ocurre, y que B también ocurre, es equivalente a la ocurrencia del suceso $A \cap B = \{5\}$ y como la información que se posee reduce el espacio muestral al suceso B , el cual contiene dos puntos igualmente posibles, entonces, según la definición clásica, la respuesta es $1/2$.

Sea el espacio de probabilidad (Ω, \mathcal{F}, P) . La probabilidad de un suceso A condicionada a la información de que otro suceso B ocurre se denomina *probabilidad condicional de A dado B* , y se denota por $P(A|B)$.

En el ejemplo que estamos analizando, $P(A|B) = 1/2$. Notemos que

$$P(A|B) = \frac{1}{2} = \frac{1/6}{2/6} = \frac{P(A \cap B)}{P(B)}.$$

De manera similar se tiene que $P(B|A) = 1/3$. El razonamiento que se acaba de hacer y la fórmula obtenida, son válidos en general para los modelos clásico y geométrico, y para la frecuencia relativa, lo cual hace natural la siguiente definición.

Definición 1.

Dados dos sucesos A y B , y $P(B) > 0$, la **probabilidad condicional de A dado B** , denotada por $P(A|B)$, se define

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Ejemplo 6.1 Una moneda equilibrada se lanza al azar en dos ocasiones. Dado que en uno de los lanzamientos apareció el escudo (suceso B), ¿cuál es la probabilidad de que en el otro lanzamiento también se haya obtenido el escudo (suceso A)?

Sean

$$\Omega = \{(0, 0), (1, 1), (0, 1), (1, 0)\}, \quad A = \{(1, 1)\}, \quad B = \{(1, 1), (0, 1), (1, 0)\}.$$

La información sobre la ocurrencia de B reduce el espacio muestral a tres puntos igualmente posibles, de los cuales uno solo es favorable a la ocurrencia de A , luego, $P(A|B) = 1/3$. Por supuesto que usando la definición de probabilidad condicional se obtiene igual resultado:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3.$$

Notemos que si el problema se plantea de manera diferente: dado que en el segundo lanzamiento apareció el escudo (suceso B ahora), ¿cuál es la probabilidad de que en el otro lanzamiento también haya aparecido el escudo? En este caso, $B = \{(0, 1), (1, 1)\}$ y $P(A|B) = 1/2$. ■

Sean el espacio de probabilidad (Ω, \mathcal{F}, P) , $B \in \mathcal{F}$ y $P(B) > 0$. Entonces $P(\Omega|B) = 1$, y si $\{A_n\}$ es una sucesión de eventos disjuntos dos a dos, se tiene que

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

es decir, se cumple la propiedad de σ -aditividad. Luego P_B , definida como $P_B(A) = P(A|B)$, para todo $A \in \mathcal{F}$, satisface la definición de probabilidad, y es por lo tanto otra probabilidad sobre el espacio medible (Ω, \mathcal{F}) , que como tal satisface todas las propiedades inherentes a una probabilidad.

De la definición de probabilidad condicional, inmediatamente se obtiene la siguiente fórmula para el cálculo de la probabilidad de ocurrencia conjunta de dos sucesos:

$$P(A \cap B) = P(A|B)P(B), \quad (6.1)$$

cuando $P(B) > 0$.

Si es posible calcular la probabilidad condicional $P(A|B)$, restringiendo el espacio muestral al evento B , entonces la fórmula 6.1 resulta de gran utilidad para calcular $P(A \cap B)$. El siguiente ejemplo ilustra esta idea.

Ejemplo 6.2 Un lote de N artículos contiene M defectuosos. Se seleccionan sucesivamente sin reposición dos artículos al azar. ¿Cuál es la probabilidad de que los dos artículos seleccionados sean defectuosos?

Sean D_1 y D_2 los sucesos que consisten, respectivamente, en que el primero y segundo artículo seleccionado es defectuoso. Entonces,

$$P(D_1 \cap D_2) = P(D_2|D_1)P(D_1) = \frac{M-1}{N-1} \frac{M}{N}.$$

Recuerde que en el Ejemplo 3.8 se obtuvo una fórmula general de la teoría combinatoria para el cálculo de la probabilidad $P_n(k; N, M)$, de que al seleccionar n artículos de entre el total de N , k de ellos resulten defectuosos. Por supuesto que usando esa fórmula se obtiene igual resultado:

$$P(D_1 \cap D_2) = P_2(2; N, M) = \frac{\binom{M}{2}}{\binom{N}{2}} = \frac{M-1}{N-1} \frac{M}{N}.$$

■

Por medio de la fórmula 6.1 no es difícil deducir la siguiente proposición.

Proposición 1.

Sean tres sucesos A_1 , A_2 y A_3 , y $P(A_1 \cap A_2) > 0$. Entonces

$$P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2)P(A_2|A_1)P(A_1).$$

Ejercicio 1.

Extienda la fórmula de la Proposición 1 al caso general de n sucesos, $n \geq 3$, y pruebe la fórmula.

6.2. Probabilidad total y fórmula de Bayes

Ejemplo 6.3 Consideremos, como en el Ejemplo 6.2, un lote de N artículos que contiene M defectuosos. Se seleccionan sin reposición dos artículos al azar. ¿Cuál es la probabilidad de que el segundo artículo seleccionado sea defectuoso?

Denotemos por D_1 y D_2 los sucesos que consisten, respectivamente, en que el primero y el segundo artículo seleccionado es defectuoso. Entonces, utilizando la aditividad finita de la probabilidad, y la fórmula para el cálculo de la probabilidad de ocurrencia conjunta de dos sucesos, se obtiene

$$\begin{aligned} P(D_2) &= P((D_1 \cap D_2) \cup (D_1^c \cap D_2)) \\ &= P(D_1 \cap D_2) + P(D_1^c \cap D_2) \\ &= P(D_2|D_1)P(D_1) + P(D_2|D_1^c)P(D_1^c) \\ &= \frac{M-1}{N-1} \frac{M}{N} + \frac{M}{N-1} \frac{N-M}{N} = \frac{M}{N}. \end{aligned}$$

En general, la probabilidad de que el i -ésimo artículo seleccionado, $1, 2, \dots, N$, sea defectuoso, es igual a M/N . Este resultado se obtiene más fácil usando la teoría combinatoria (vea el Ejercicio 8 de la Sección 3.4). Intuitivamente el resultado puede parecer no cierto.

■

El siguiente teorema constituye una generalización de la fórmula obtenida en el ejemplo anterior.

Teorema 2 (Fórmula de la Probabilidad Total).

Sean los sucesos B_1, B_2, \dots, B_n tales que $\bigcup_{i=1}^n B_i = \Omega$, $B_i \cap B_j = \emptyset$, $i \neq j$ y $P(B_i) > 0$, $1 \leq i \leq n$, y sea A otro suceso. Entonces

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

Demostración:

Los conjuntos $A \cap B_i$, $1 \leq i \leq n$, son disjuntos. Entonces, usando la aditividad finita de la probabilidad y la fórmula para el cálculo de la probabilidad de ocurrencia conjunta de dos sucesos, se halla que

$$\begin{aligned} P(A) &= P\left(A \cap \bigcup_{i=1}^n B_i\right) = P\left(\bigcup_{i=1}^n A \cap B_i\right) \\ &= \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i)P(B_i). \end{aligned}$$

■

Teorema 3 (Fórmula de Bayes I).

Sean A y B sucesos. Si $P(A) > 0$ y $P(B) > 0$, entonces

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Demostración:

El resultado es evidente de las igualdades

$$P(A \cap B) = P(A|B)P(B) \quad \text{y} \quad P(A \cap B) = P(B|A)P(A).$$

■

De los dos teoremas anteriores inmediatamente se deduce el siguiente:

Teorema 4 (Fórmula de Bayes II).

En las condiciones que tiene lugar la fórmula de la probabilidad total, Teorema 2, si además $P(A) > 0$, se tiene

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}, \quad 1 \leq i \leq n.$$

Observación 2 La fórmula de la probabilidad total y el teorema de Bayes también tienen lugar, si en lugar de considerar una partición finita de Ω en eventos se considera una partición infinita numerable $\{B_i\}$. Note que $\sum_{i=1}^{\infty} P(B_i|A) = 1$. ▲

Ejemplo 6.4 Supongamos que en las condiciones del Ejemplo 6.3, el segundo artículo que se extrae del lote es defectuoso, pero se desconoce cuál fue el resultado de la primera elección. ¿Cuál es la probabilidad de que el primer artículo haya sido también defectuoso?

Usando la fórmula de la probabilidad total, en el Ejemplo 6.3 se halló que $P(D_2) = M/N$. Entonces por la fórmula de Bayes se obtiene la respuesta a la pregunta formulada:

$$P(D_1|D_2) = \frac{P(D_2|D_1)P(D_1)}{P(D_2)} = \frac{\frac{M-1}{N-1} \frac{M}{N}}{\frac{M}{N}} = \frac{M-1}{N-1}.$$

Además, la probabilidad de que el primer artículo no haya sido defectuoso, dado que el segundo lo fue, se halla

$$P(D_1^c|D_2) = 1 - P(D_1|D_2) = \frac{N-M}{N-1}.$$

Es interesante notar que, en particular, en este ejemplo se cumple que:

$$P(D_1|D_2) = P(D_2|D_1) \quad \text{y} \quad P(D_1^c|D_2) = P(D_2^c|D_1).$$

Ejemplo 6.5 *Problema del caminante.* (Vea la Figura 6.1) Considere un caminante que parte de un punto O, desde el cual puede llegar por tres caminos diferentes, y por uno y solo uno, a los puntos 1, 2 y 3. Desde 1 existe un único camino, y este conduce al punto F; desde 2 existe un camino que conduce a

F y otro que no; y desde 3 existe un único camino que conduce a F y otros dos que no. Supongamos que el caminante selecciona siempre al azar cada camino. Calculemos la probabilidad de que, partiendo de O , el caminante:

- Llegue a F .
- Dado que llegó a F , esto haya sido pasando por el punto 2.

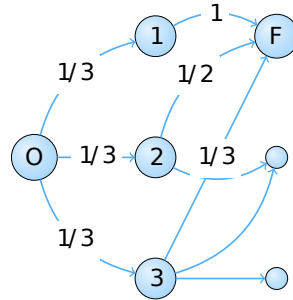


Figura 6.1: Posibles recorridos del caminante

Denotemos por A el suceso que consiste en que el caminante llega a F partiendo de O , y por B_i el suceso que consiste en que el caminante pasa por el punto i , $i = 1, 2, 3$. Entonces $P(B_i) = 1/3$, $i = 1, 2, 3$, y

$$\text{a) } P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

$$= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{11}{18}.$$

$$\text{b) } P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{1/2 \cdot 1/3}{11/18} = \frac{3}{11} \approx 0,272.$$

De manera análoga se calculan $P(B_1|A) = 6/11$ y $P(B_3|A) = 2/11$. Luego,

$$P(B_1|A) > P(B_2|A) > P(B_3|A).$$

Si el interés consiste solo en el orden de las probabilidades $P(B_i|A)$, $i = 1, 2, 3$, y no en sus valores exactos, entonces, como las fórmulas de Bayes para estas probabilidades tienen el denominador común $P(A)$, es suficiente comparar los numeradores de las fórmulas correspondientes a cada probabilidad.

Si ahora tomamos, $P(B_1) = P(B_2) = 1/6$ y $P(B_3) = 4/6$, es fácil hallar que

$$P(B_3|A) > P(B_1|A) > P(B_2|A),$$

resultado diferente del caso original, en el cual las probabilidades $P(B_i)$, $i = 1, 2, 3$, se asumieron iguales. ■

En la fórmula de Bayes las probabilidades $P(B_i)$, $i \geq 1$, se calculan sin información previa sobre la ocurrencia del evento A , y las probabilidades $P(B_i|A)$, $i \geq 1$, se calculan con la información posterior de que A ocurrió u ocurre. Por eso las primeras se denominan *probabilidades a priori* y las segundas *probabilidades a posteriori*.

Observación 3 Es importante señalar que, como se ha visto en los ejemplos recientes, en la solución de muchos problemas de carácter práctico, habitualmente no se describen el espacio muestral y los sucesos de interés en términos de los conjuntos que los representan, sino solo se describen de manera cualitativa. ▲

6.3. Independencia de sucesos

Intuitivamente, un suceso A es independiente de un suceso B si la ocurrencia de B no influye en la probabilidad de ocurrencia de A , es decir, si $P(A|B) = P(A)$ cuando $P(B) > 0$. Entonces, de la fórmula

$$P(A \cap B) = P(A|B)P(B),$$

se deduce que

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

No es difícil comprobar que cuando $P(A) > 0$ y $P(B) > 0$, si A es independiente de B , entonces B también es independiente de A ; por eso es natural decir que A y B son independientes. Esta idea intuitiva de independencia motiva la siguiente definición.

Definición 2.

Los sucesos A y B se dice que son **independientes**, si

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

Cuando la igualdad anterior no se cumple, se dice que los sucesos son **dependientes**.

No es complicado verificar que todo suceso A es independiente de sí mismo si, y solo si, $P(A) = 0$ o $P(A) = 1$, y además, que si $P(A) = 0$ o $P(A) = 1$, entonces A y un suceso cualquiera B son independientes; en particular, B y Ω son independientes, y también B y \emptyset .

Sean A y B dos sucesos tales que $P(A) > 0$ y $P(B) > 0$. Entonces si A y B son excluyentes, son dependientes.

En la teoría de las probabilidades los conceptos de independencia y dependencia pueden modelar situaciones en las cuales no se trata de una independencia o dependencia *causal*. Estos conceptos probabilísticos son más amplios de lo que usualmente sugieren la experiencia y la intuición.

Ejemplo 6.6 Consideremos en el experimento del lanzamiento al azar de un dado equilibrado los sucesos

$$A = \{3, 6\}, \quad B = \{3, 5\}, \quad C = \{2, 3, 5\}.$$

Los sucesos A y B son dependientes, y también son dependientes B y C ; sin embargo, A y C son independientes, lo cual podría parecer no natural.

Si ahora se truca el dado de manera que cuando se lance al azar solo pueden aparecer, con igual probabilidad $1/3$, los números 1, 2 ó 6, entonces A y B son independientes, y B y C también lo son, pero A y C son dependientes.

■

El ejemplo anterior muestra también, que dos sucesos (A y B en el ejemplo) pueden ser dependientes de acuerdo a una medida de probabilidad, e independientes de acuerdo a otra medida de probabilidad.

Teorema 5.

Sean A y B sucesos. Las siguientes afirmaciones son equivalentes:

1. A y B son independientes
2. A y B^c son independientes
3. A^c y B^c son independientes

Demostración:

Probemos que 1 implica 2. Efectivamente, si A y B son independientes, en-

tonces

$$\begin{aligned} P(A \cap B^c) &= P(A \setminus (A \cap B)) = P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] = P(A)P(B^c), \end{aligned}$$

luego A y B^c son independientes. Hemos probado que si dos sucesos son independientes, entonces también lo son uno de ellos y el contrario del otro. De esta observación inmediatamente se deducen las restantes implicaciones. ■

Definición 3.

Tres sucesos, A , B y C , se dice que son **independientes** (o *conjuntamente independientes*), si son independientes dos a dos y, además,

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

Como muestra el siguiente ejemplo, tres sucesos pueden ser independientes dos a dos y no ser conjuntamente independientes.

Ejemplo 6.7 Sean $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $P(\omega_i) = 1/4$, $i = 0, 1, 2, 3$, y los sucesos $A_i = \{\omega_0, \omega_i\}$, $i = 1, 2, 3$. Se tiene que

$$P(A_i \cap A_j) = P(\omega_0) = 1/4 = P(A_i)P(A_j), \quad i \neq j.$$

Pero,

$$P(A_1 \cap A_2 \cap A_3) = P(\omega_0) = 1/4 \neq P(A_1)P(A_2)P(A_3) = 1/8. \quad \blacksquare$$

La definición de independencia de tres sucesos se extiende de manera natural para un número finito de sucesos.

Definición 4.

Los sucesos A_1, A_2, \dots, A_n se dice que son **independientes**, si para todo $k = 2, 3, \dots, n$ y para todos $i_1, i_2, \dots, i_k, 1 \leq i_1 < \dots < i_k \leq n$, se cumple

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

No es difícil comprender que para una colección de sucesos conjuntamente independientes se cumple la siguiente propiedad: cualquier combinación, en el sentido de las operaciones con conjuntos, de un grupo de sucesos, es independiente de cualquier combinación de otro grupo de sucesos diferente. Por ejemplo, si los sucesos A, B, C y D son conjuntamente independientes, entonces los sucesos $A \cup B$ y $C \cup D$, y también los sucesos $A \cup B$, y $C \cap D$ son independientes.

Para una sucesión infinita de sucesos se tiene la siguiente definición de independencia.

Definición 5.

Los sucesos A_1, A_2, \dots son *independientes*, si para todo $k = 2, 3, \dots$, y para todos $i_1, i_2, \dots, i_k, 1 \leq i_1 < i_2 < \dots < i_k$, se cumple

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}).$$

Consideremos un espacio medible (Ω, \mathcal{F}) . Una σ -álgebra \mathcal{F}' en Ω se dice que es una **sub σ -álgebra** de la σ -álgebra \mathcal{F} si $\mathcal{F}' \subset \mathcal{F}$.

Definición 6.

Las sub σ -álgebras $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ de \mathcal{F} se dice que son **σ -álgebras independientes**, si para todos $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$ los sucesos A_1, A_2, \dots, A_n son independientes.

Ejemplo 6.8 Sean los sucesos A_1 y A_2 . Probemos que

$$\mathcal{F}_1 = \{A_1, A_1^c, \emptyset, \Omega\} \quad \text{y} \quad \mathcal{F}_2 = \{A_2, A_2^c, \emptyset, \Omega\}$$

son σ -álgebras independientes si, y solo si, A_1 y A_2 son independientes.

El cumplimiento de la parte *solo si* es evidente. Veamos la parte *si*: \emptyset y Ω son independientes de todo suceso; por otra parte, si A_1 y A_2 son independientes, entonces, por el Teorema 6.3, también son independientes A_1 y A_2^c , A_1^c y A_2 , A_1^c y A_2^c . ■

Observación 4 Es importante destacar que muchas veces, tanto en el trabajo teórico, como en la modelación de problemas prácticos, la independencia no se demuestra, sino se asume como parte integrante del modelo. ▲

6.4. Lema de Borel-Cantelli. Parte II

Teorema 6 (Lema de Borel-Cantelli- Parte II).

Sea $\{A_n\}$ una sucesión de eventos independientes. Se cumple que:

$$\text{Si } \sum_{n=1}^{\infty} P(A_n) = \infty, \text{ entonces } P(A_n \text{ i.f.}) = 1.$$

Demostración:

Debemos probar que

$$1 - P(A_n \text{ i.f.}) = P\left(\left(\limsup_n A_n\right)^c\right) = 0.$$

Para ello, como

$$P\left(\left(\limsup_n A_n\right)^c\right) = P\left(\liminf_n A_n^c\right) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right),$$

usando la sub-aditividad de la probabilidad es suficiente probar que para todo $n \geq 1$ se cumple

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0. \quad (6.2)$$

Si los sucesos A_1, A_2, \dots son independientes, entonces también son independientes los sucesos A_1^c, A_2^c, \dots . Luego para todo $m \geq n$ se tiene

$$P\left(\bigcap_{k=n}^m A_k^c\right) = \prod_{k=n}^m P(A_k^c).$$

Aplicando logaritmo en la igualdad anterior, y utilizando la desigualdad $\log(1-x) \leq -x, 0 \leq x < 1$, se halla que

$$\log\left(P\left(\bigcap_{k=n}^m A_k^c\right)\right) = \sum_{k=n}^m \log(1 - P(A_k)) \leq -\sum_{k=n}^m P(A_k),$$

de donde, como la serie $\sum_{n=1}^{\infty} P(A_n)$ es divergente, se obtiene

$$\lim_m \log\left(P\left(\bigcap_{k=n}^m A_k^c\right)\right) = -\infty,$$

y entonces

$$\lim_m P\left(\bigcap_{k=n}^m A_k^c\right) = 0.$$

Por lo tanto, usando la continuidad de la probabilidad, finalmente se demuestra (6.2):

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = P\left(\lim_m \bigcap_{k=n}^m A_k^c\right) = \lim_m P\left(\bigcap_{k=n}^m A_k^c\right) = 0,$$

para todo $n \geq 1$. ■

Si en el teorema anterior se elimina la suposición de independencia de los sucesos, entonces la afirmación que se establece en el mismo podría no cumplirse. Por ejemplo, los sucesos $A_i = A_1, i \geq 2$, son dependientes si $0 < P(A_1) < 1$, y además se cumple que $\sum_{n=1}^{\infty} P(A_n) = \infty$; sin embargo, $P(\limsup_n A_n) = P(A_1) \neq 1$.

De las Partes I y II del Lema de Borel-Cantelli inmediatamente se deduce que si $\{A_n\}$ es una sucesión de eventos independientes, entonces solo existen dos posibles valores para $P(A_n \text{ i.f.})$. Esta probabilidad es igual a 0 o a 1.

6.5. Problemas y ejercicios

1. Sean A y B sucesos, y $P(A)P(B) > 0$. Pruebe que entonces $P(A|B) \geq P(A)$ si, y solo si, $P(B|A) \geq P(B)$.
2. Considere tres cartas. Una carta es roja por ambas caras, otra carta es negra por ambas caras, y la tercera es roja por una cara y negra por la otra. Se selecciona una carta al azar, de la cual solo es visible una cara, que resulta negra. ¿Cuál es la probabilidad de que la otra cara de la carta seleccionada también sea negra? **R:** $2/3$.
3. **Problema de Monty Hall*. Considere tres bolsas, una de las cuales contiene un premio. Usted selecciona una de las bolsas al azar, y luego, de las dos restantes, le muestran una bolsa vacía. A continuación puede seguir una de las dos siguientes estrategias:
 - i. Quedarse con la bolsa seleccionada al azar.
 - ii. Seleccionar la otra bolsa que queda, de contenido desconocido.
 - a) Calcule la probabilidad de obtener el premio con cada una de las estrategias.
 - b) Implemente un programa de simulación en lenguaje Python que permita calcular aproximadamente la probabilidad de obtener el premio usando la estrategia ii.
4. Una prueba sobre una enfermedad da negativo si no existe la enfermedad, y positivo si esta existe, con probabilidad 0,95 en ambos casos. Se conoce que en la población una persona de cada 2000 tiene la enfermedad.
 - a) Si la prueba le da positivo a una persona, ¿cuál es la probabilidad de que realmente tenga la enfermedad?
 - b) Si le da negativo, ¿cuál es la probabilidad de que realmente no esté enferma?**R:** a) $19/2018 \approx 0,01$.
5. Se lanzan al azar dos dados equilibrados. Verifique que la probabilidad de que la suma de los puntos sea 7 es independiente del resultado del primer lanzamiento.

6. Suponga que de los números 2, 3, 5, 30 se selecciona uno al azar. Sea A_k el suceso que consiste en que el número seleccionado es divisible por k , $k = 2, 3, 5$. Analice la independencia dos a dos de los tres sucesos y su independencia conjunta.
7. Demuestre que la independencia conjunta de los sucesos A , B y C implica la independencia de los sucesos:
 - a) A y $B \cup C$;
 - b) A y $B \cap C$;
 - c) A^c , B^c y C^c .
8. Suponga que los sucesos A y B_1 son independientes y que también lo son los sucesos A y B_2 . Pruebe que entonces A y $B_1 \cup B_2$ son independientes si, y solo si, A y $B_1 \cap B_2$ son independientes.
9. * Sean $\{A_n\}$ y $\{B_n\}$ dos sucesiones crecientes de sucesos con respectivos límites A y B . Pruebe que si A_n y B_n son independientes para cada $n \geq 1$, entonces A y B son independientes.
10. Pruebe que la probabilidad de que ninguno de los sucesos independientes A_1, A_2, \dots, A_n ocurra, es menor o igual que $\exp(-\sum_{i=1}^n P(A_i))$.

ESPACIO PRODUCTO

En este capítulo se considera el espacio producto asociado al caso especial de repetición de experimentos con un número finito de resultados. Se establece la relación entre la noción intuitiva de experimentos independientes y el espacio producto.

7.1. Espacio producto

Consideremos n espacios de probabilidad discretos,

$$(\Omega_1, \mathcal{P}(\Omega_1), P_1), (\Omega_2, \mathcal{P}(\Omega_2), P_2), \dots, (\Omega_n, \mathcal{P}(\Omega_n), P_n),$$

donde, recordemos, $\mathcal{P}(\Omega_i)$ denota el conjunto potencia de Ω_i , $i = 1, 2, \dots, n$. Construyamos un espacio de probabilidad en el cual se puedan representar sucesos de cada uno de estos espacios de tal manera tal que estos sean independientes.

Sea el espacio muestral $\Omega^n = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$. Definamos para todo $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$ la probabilidad P^n como

$$P^n(\omega) = P_1(\omega_1)P_2(\omega_2) \cdots P_n(\omega_n).$$

Sea $A_i \subset \Omega_i, i = 1, 2, \dots, n$. Observe que

$$\begin{aligned} P^n(A_1 \times A_2 \times \dots \times A_n) &= \sum_{\omega \in A_1 \times A_2 \times \dots \times A_n} P^n(\omega) \\ &= \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} \dots \sum_{\omega_n \in A_n} P_1(\omega_1) P_2(\omega_2) \dots P_n(\omega_n) \\ &= P_1(A_1) P_2(A_2) \dots P_n(A_n). \end{aligned} \quad (7.1)$$

Puesto que $P^n(\omega) \geq 0$, para todo $\omega \in \Omega^n$, y de (7.1) se obtiene que $P^n(\Omega^n) = 1$, entonces P^n es, efectivamente, una probabilidad definida sobre $\mathcal{P}(\Omega^n)$.

El espacio de probabilidad discreto $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$, así definido, se denomina **espacio de probabilidad producto** de los espacios $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$, $i = 1, 2, \dots, n$.

Sean ahora,

$$\begin{aligned} \tilde{A}_1 &= A_1 \times \Omega_2 \times \dots \times \Omega_{n-1} \times \Omega_n, \\ \tilde{A}_i &= \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n, \quad i = 2, 3, \dots, n-1, \\ \tilde{A}_n &= \Omega_1 \times \Omega_2 \times \dots \times \Omega_{n-1} \times A_n. \end{aligned}$$

Utilizando (7.1) se deduce que

$$P^n(\tilde{A}_i) = P_i(A_i), \quad i = 1, 2, \dots, n. \quad (7.2)$$

Por consiguiente, el suceso \tilde{A}_i constituye la representación en el espacio producto $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ del suceso $A_i, i = 1, 2, \dots, n$.

Teorema 1.

En el espacio producto $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$, los sucesos $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ son conjuntamente independientes.

Demostración:

Probemos la independencia de dos sucesos, digamos, \tilde{A}_1 y \tilde{A}_2 . En efecto, se cumple que

$$P^n(\tilde{A}_1 \cap \tilde{A}_2) = P^n(A_1 \times A_2 \times \Omega_3 \times \dots \times \Omega_n) = P_1(A_1) P_2(A_2) = P^n(\tilde{A}_1) P^n(\tilde{A}_2),$$

donde las dos últimas igualdades se deducen usando (7.1) y (7.2), respectivamente. El caso general de independencia conjunta se prueba de forma análoga. ■

Por supuesto que otros sucesos del espacio producto, diferentes de $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$, no necesariamente son independientes. Por ejemplo, en general los sucesos \tilde{A}_1 y $\tilde{A}_1 \cup \tilde{A}_2$ no son independientes.

Ejemplo 7.1 Consideremos un sistema constituido por 3 componentes. El sistema falla si fallan al menos dos de sus componentes. Supongamos que inicialmente los tres componentes funcionan y que la probabilidad de que el componente i este funcionando al cabo del tiempo t es igual a p_i , $i = 1, 2, 3$. Asumamos que los fallos de los componentes son independientes entre sí. Hallemos la probabilidad de que el sistema no haya fallado al cabo del tiempo t .

Sea A_i el suceso que consiste en que en el instante t funciona (éxito) el componente i , $i = 1, 2, 3$. Denotemos $q_i = 1 - p_i$, $i = 1, 2, 3$. Tomemos, $\Omega = \{0, 1\}$, $A_i = \{1\}$, $P_i(\omega_i) = p_i^{\omega_i} q_i^{1-\omega_i}$, $\omega_i \in \Omega_i$, $i = 1, 2, 3$. En el espacio producto de los espacios $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$, $i = 1, 2, 3$, los sucesos $\tilde{A}_i, i = 1, 2, 3$, son independientes.

El sistema funciona en el instante t cuando al cabo de ese tiempo ninguno de los componente ha fallado o solo uno de ellos ha fallado, lo cual se representa por el suceso,

$$(\tilde{A}_1 \cap \tilde{A}_2 \cap \tilde{A}_3) \cup (\tilde{A}_1^c \cap \tilde{A}_2 \cap \tilde{A}_3) \cup (\tilde{A}_1 \cap \tilde{A}_2^c \cap \tilde{A}_3) \cup (\tilde{A}_1 \cap \tilde{A}_2 \cap \tilde{A}_3^c),$$

cuya probabilidad es igual a

$$p_1 p_2 p_3 + q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3.$$

■

7.2. Modelo de Bernoulli

Consideremos una serie de n experimentos, en cada uno de los cuales interesa solo la ocurrencia o no de cierto suceso, cuya probabilidad de ocurrir en cada experimento es igual a p , $0 < p < 1$. Por medio de un espacio producto, construiremos un modelo que permita considerar los experimentos de esta serie independientes.

Sea $\Omega_i = \{0, 1\}$ el espacio muestral correspondiente al experimento i , $i = 1, 2, \dots, n$, donde, como es habitual, por 1 se representa la ocurrencia del suceso de interés o éxito, y por 0 su no ocurrencia o fracaso. Definamos

$$P_i(\omega_i) = p^{\omega_i} q^{1-\omega_i}, \quad \text{si } \omega_i \in \Omega_i, \quad i = 1, 2, \dots, n,$$

donde $q = 1 - p$.

Denotemos por A_i el suceso que representa el éxito en el experimento i , es decir, $A_i = \{1\}$; entonces se tiene que $P_i(A_i) = p$ y la σ -álgebra generada por el conjunto A_i es

$$\sigma(\{A_i\}) = \{\Omega_i, \emptyset, A_i, A_i^c\} = \mathcal{P}(\Omega_i), \quad i = 1, 2, \dots, n.$$

Por conveniencia en lo que sigue, denotaremos $\mathcal{F}_i = \sigma(\{A_i\})$. El espacio de probabilidad $(\Omega_i, \mathcal{F}_i, P_i)$ está asociado al experimento i , $i = 1, 2, \dots, n$. Construyamos el espacio producto de estos espacios de probabilidad.

Sea $\Omega^n = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$. Para todo $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega^n$, definamos

$$P^n(\omega) = P_1(\omega_1)P_2(\omega_2) \cdots P_n(\omega_n) = p^{\sum_{i=1}^n \omega_i} q^{n - \sum_{i=1}^n \omega_i}.$$

El espacio de probabilidad $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ es el espacio producto.

El suceso

$$\tilde{A}_i = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega^n : \omega_i = 1\}$$

representa, en el espacio producto, el éxito en el experimento i , además, $P^n(\tilde{A}_i) = p$, $i = 1, 2, \dots, n$.

Por otro lado, la σ -álgebra $\tilde{\mathcal{F}}_i = \{\Omega^n, \emptyset, \tilde{A}_i, \tilde{A}_i^c\}$ solo contiene información sobre la ocurrencia o no de éxito en el experimento i . Por consiguiente, en el espacio producto asociamos la σ -álgebra $\tilde{\mathcal{F}}_i$ con un experimento en el cual puede ocurrir el suceso \tilde{A}_i con probabilidad p , $i = 1, 2, \dots, n$.

Por el Teorema 1, los sucesos $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_n$ son conjuntamente independientes, así que también lo son las σ -álgebras $\tilde{\mathcal{F}}_1, \dots, \tilde{\mathcal{F}}_n$. Como cada una de estas σ -álgebras está asociada a un experimento, entonces se dice que la serie de n experimentos es una serie de **experimentos independientes**.

El modelo que hemos descrito se denomina **modelo de Bernoulli**, y a cada uno de los experimentos de la serie de n experimentos se le llama *experimento (o prueba) de Bernoulli*.

Denotemos por $P_n(k; p)$ la probabilidad de que en exactamente k de los n experimentos ocurra éxito. Entonces

$$\begin{aligned}
 P_n(k; p) &= P^n \left(\left\{ \omega \in \Omega^n : \sum_{i=1}^n \omega_i = k \right\} \right) \\
 &= \sum_{\omega: \sum_{i=1}^n \omega_i = k} P^n(\omega) \\
 &= \sum_{\omega: \sum_{i=1}^n \omega_i = k} p^{\sum_{i=1}^n \omega_i} q^{n - \sum_{i=1}^n \omega_i} \\
 &= \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n.
 \end{aligned}$$

La última igualdad tiene lugar debido a que la suma $\sum_{i=1}^n \omega_i$ es igual a k en $\binom{n}{k}$ formas diferentes, y a cada una ellas le corresponde la probabilidad $p^k q^{n-k}$, $k = 0, 1, \dots, n$.

Observe que

$$\sum_{k=0}^n P_n(k; p) = 1.$$

Note además, que la probabilidad $P_n(k; p)$ es el coeficiente de x^k en el desarrollo de $(px + q)^n$ en potencias de x . Por eso al modelo de Bernoulli también se le llama **modelo binomial**.

Desde el punto de vista práctico, el modelo de Bernoulli se asume cuando se considera una serie de n experimentos que pueden suponerse independientes, en cada uno de los cuales puede ocurrir con probabilidad p cierto suceso de interés.

Ejemplo 7.2 Si en el Ejemplo 7.1 tomamos $p_i = p$, $i = 1, 2, 3$, $q = 1 - p$. Entonces, con las notaciones correspondientes al modelo de Bernoulli, la probabilidad de que el sistema esté funcionando al cabo del tiempo t se calcula de la siguiente forma,

$$P_3(0; q) + P_3(1; q) = \binom{3}{0} q^0 p^3 + \binom{3}{1} qp^2 = p^3 + 3qp^2.$$

■

Ejemplo 7.3 Se tienen dos monedas idénticas en su aspecto físico; sin embargo, se conoce que cuando se lanzan al azar, las probabilidades de aparición

de escudo para cada una de las monedas, que denominaremos I y II , son $1/2$ y $2/3$, respectivamente. Se lanza 100 veces al azar una de las monedas y se debe decidir cuál de las monedas fue lanzada. Asumamos la siguiente regla de decisión: si el escudo aparece más de 60 veces, la moneda lanzada es la II , en caso contrario la moneda lanzada es la I .

En este problema se pueden calcular las probabilidades de los dos errores posibles al utilizar la regla de decisión: el primer error consiste en decidir que la moneda lanzada fue la II , habiendo sido en realidad la I ; y el segundo, decidir que la moneda lanzada fue la I , si en realidad fue la II . Denotemos por α y β las probabilidades del primero y segundo error, respectivamente. Luego α es la probabilidad de que al lanzar 100 veces la moneda equilibrada, en más de 60 ocasiones aparezca el escudo, y β la probabilidad de que si se lanza 100 veces la otra moneda, en 60 o menos ocasiones aparezca el escudo. Entonces

$$\alpha = \sum_{k=61}^{100} P_{100}(k; 1/2) = \sum_{k=61}^{100} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k} = 0,018,$$

y

$$\beta = \sum_{k=0}^{60} P_{100}(k; 2/3) = \sum_{k=0}^{60} \binom{100}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{100-k} = 0,097.$$

Los valores de α y β se calcularon aproximadamente con redondeo en la tercera cifra decimal.

A medida que aumenta el número de lanzamientos de la moneda se pueden seleccionar reglas de decisión con valores de α y de β cada vez menores.

El siguiente fragmento de código Python muestra como calcular los valores de α y β .

```
from scipy.stats import binom
alfa=1-binom.cdf(60, 100, 1/2)
beta=binom.cdf(60, 100, 2/3)
```

■

Ejemplo 7.4 Consideremos el *Problema del caminante*, Ejemplo 6.5. Supongamos ahora que hay tres caminantes, 1, 2 y 3, y que se selecciona al azar uno de ellos, el cual partirá de O en tres ocasiones. Se conoce que el caminante i siempre pasa por el punto i , $i = 1, 2, 3$. Si el caminante seleccionado llegó a F en las tres ocasiones, ¿cuál es la probabilidad de que haya sido el caminante 2?

Denotemos por A el suceso que consiste en que el caminante llegó a F en las tres ocasiones en que partió de 0. Las tres posibles llegadas a F son independientes entre sí. Denotemos por B_i el suceso que consiste en que el caminante seleccionado es el i , $i = 1, 2, 3$. Entonces $P(B_i) = 1/3$, $i = 1, 2, 3$. Por otro lado,

$$P(A|B_1) = 1, \quad P(A|B_2) = (1/2)^3, \quad P(A|B_3) = (1/3)^3.$$

Luego,

$$P(B_2|A) = \frac{(1/2)^3 \cdot 1/3}{1/3 + (1/2)^3 \cdot 1/3 + (1/3)^3 \cdot 1/3} = 27/251 \approx 0,107.$$

■

Ejemplo 7.5 Sea un sistema constituido por dos componentes, denotados por 1 y 2, que funciona solo si funcionan ambos componentes. Las probabilidades de funcionamiento hasta el instante t (éxito) de cada componente son p_1 y p_2 , respectivamente, $0 < p_1, p_2 < 1$. Los fallos de los componentes se asumen independientes entre sí. Inicialmente ambos componentes funcionan, pero al cabo del tiempo t se detectó que el sistema falló.

Observe que se consideran dos pruebas independientes asociadas al funcionamiento de cada componente, y cuando las probabilidades de éxito p_1 y p_2 son diferentes no corresponde al modelo de Bernoulli.

a) ¿Qué es más probable, que hayan fallado los dos componentes, o que haya fallado solo el componente 1?

b) Supongamos que $p_1 = p_2 = p$. ¿Qué es más probable, que hayan fallado los dos componentes, o que haya fallado solo uno de los dos?

Para responder las preguntas planteadas usaremos la fórmula de Bayes, pero no será necesario calcular su denominador. Cuando se detecta el fallo en el instante t , es posible que haya ocurrido uno de los siguientes sucesos mutuamente excluyentes:

B_1 - falló el componente 1 y no falló el 2,

B_2 - falló el componente 2 y no falló el 1,

B_3 - falló el componente 1 y falló el 2,

Estos tres sucesos, y el suceso B_4 - no falló el componente 1 y no falló el 2, son mutuamente excluyentes y siempre ocurre uno de ellos, pero el fallo del sistema solo puede ocurrir con uno de los tres primeros. Por la independencia

de los fallos de los componentes se halla que

$$P(B_1) = (1 - p_1)p_2, \quad P(B_2) = (1 - p_2)p_1,$$

$$P(B_3) = (1 - p_1)(1 - p_2), \quad P(B_4) = p_1p_2.$$

Por otro lado, denotando por F el fallo del sistema se tiene,

$$P(F|B_1) = P(F|B_2) = P(F|B_3) = 1 \quad \text{y} \quad P(F|B_4) = 0.$$

De donde, usando la fórmula de Bayes, tenemos que

$$P(B_4|F) = 0,$$

$$P(B_1|F) \propto P(F|B_1)P(B_1) = (1 - p_1)p_2,$$

$$P(B_2|F) \propto P(F|B_2)P(B_2) = (1 - p_2)p_1,$$

$$P(B_3|F) \propto P(F|B_3)P(B_3) = (1 - p_1)(1 - p_2),$$

donde la constante de proporcionalidad es igual a $1/P(F)$.

a) La desigualdad $P(B_3|F) > P(B_1|F)$ se cumple si, y solo si, $p_2 < 1/2$. Por consiguiente, si $p_2 < 1/2$, la probabilidad de que hayan fallado los dos componentes es mayor que la probabilidad de que haya fallado solo el componente 1, y si $p_2 > 1/2$ la primera probabilidad es menor que la segunda.

b) Como $p_1 = p_2 = p$, la desigualdad $P(B_3|F) > P((B_1 \cup B_2)|F)$ es equivalente a $(1 - p)(1 - p) > 2(1 - p)p$. Luego, si $p < 1/3$, la probabilidad de que hayan fallado los dos componentes es mayor que la probabilidad de que haya fallado solo uno de los dos, y es menor en el caso $p > 1/3$. ■

7.3. Modelo multinomial

El modelo de Bernoulli se generaliza de forma natural para experimentos independientes, en los cuales es de interés la ocurrencia en cada uno de ellos de algún suceso de una colección de más de dos sucesos excluyentes. Por ejemplo, en el lanzamiento al azar n veces de un dado, se pueden considerar los sucesos A_i , B_i y C_i que representan, respectivamente, la aparición en el lanzamiento i , $i = 1, 2, \dots, n$, de un número mayor que 3, del número 3, y de un número menor que 3, cuyas probabilidades correspondientes, en el caso de un dado equilibrado, son $3/6 = 1/2$, $1/6$ y $2/6 = 1/3$.

Formularemos este modelo sin exponer todos los detalles. Consideremos n experimentos, en cada uno de los cuales puede ocurrir uno de r , $r \geq 2$, sucesos excluyentes. La ocurrencia de cada uno de estos sucesos la asociaremos

7.4. Sucesión infinita de experimentos independientes

con los números $1, 2, \dots, r$, respectivamente. Sean

$$\Omega_i = \{1, 2, \dots, r\} \quad \Omega^n = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n.$$

$$P_i(j) = p_j, \quad j \in \Omega_i, \quad i = 1, 2, \dots, n, \quad \sum_{j=1}^r p_j = 1.$$

Supongamos que un punto muestral dado ω de Ω^n contiene exactamente k_j componentes iguales a j , $k_j = 0, 1, \dots, n$, $j = 1, 2, \dots, r$, con $\sum_{j=1}^r k_j = n$, $r \geq 2$. Entonces en el modelo de experimentos independientes se tiene que

$$P^n(\omega) = p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

El número de puntos muestrales de Ω^n con estas características es igual a

$$\frac{n!}{k_1! k_2! \dots k_r!},$$

por consiguiente, la probabilidad $P_n(k_1, k_2, \dots, k_r; p_1, p_2, \dots, p_r)$, de que en los n experimentos ocurran exactamente k_j eventos de tipo j , $j = 1, 2, \dots, r$, se calcula por la fórmula

$$P_n(k_1, k_2, \dots, k_r; p_1, p_2, \dots, p_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

Esta probabilidad es el coeficiente de $x_1^{k_1} x_2^{k_2} \dots x_r^{k_r}$ en el polinomio que se obtiene en el desarrollo de $(p_1 x_1 + p_2 x_2 + \dots + p_r x_r)^n$. A este modelo se le llama **modelo multinomial**.

7.4. Sucesión infinita de experimentos independientes

Si el conjunto finito Ω_i es el espacio muestral asociado a un experimento i , $i \geq 1$, entonces el espacio muestral correspondiente a la sucesión infinita de estos experimentos es el conjunto infinito no numerable

$$\Omega^\infty = \Omega_1 \times \Omega_2 \times \dots.$$

No es posible tratar este caso de la misma manera que cuando se considera un número finito de experimentos.

Introduciremos de manera intuitiva la noción de sucesión infinita de experimentos independientes, utilizando para ello el modelo binomial. El caso

correspondiente al modelo multinomial es análogo. En la Sección 8.3 del Capítulo 8, Complementos, se hace un examen riguroso de este tema.

Consideremos una sucesión infinita de experimentos, en cada uno de los cuales la probabilidad de éxito es igual a p , $0 < p < 1$. Denotemos $q = 1 - p$. Sea $\Omega_i = \{0, 1\}$, $i \geq 1$. El modelo de Bernoulli para una serie de n experimentos independientes es el espacio de probabilidad $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ definido en la Sección 7.2.

Sea el punto muestral $(a_1, a_2, \dots, a_n) \in \Omega^n$. El suceso que consiste en la ocurrencia consecutiva de los resultados a_1, a_2, \dots, a_n , en los primeros n experimentos de la sucesión infinita de experimentos, se representa por el conjunto

$$D_n = \{(a_1, a_2, \dots, a_n)\} \times \Omega_{n+1} \times \Omega_{n+2} \times \dots,$$

donde $a_i \in \{0, 1\}$, $i = 1, 2, \dots, n$. Denotaremos por P la medida de probabilidad correspondiente al modelo de una sucesión infinita de experimentos. Para una serie de n experimentos independientes se tiene que

$$P^n((a_1, a_2, \dots, a_n)) = p^{\sum_{i=1}^n a_i} q^{n - \sum_{i=1}^n a_i}.$$

Luego, para que los experimentos de la sucesión infinita sean independientes, la intuición indica que debe tomarse

$$P(D_n) = p^{\sum_{i=1}^n a_i} q^{n - \sum_{i=1}^n a_i}.$$

Por otro lado, si imaginamos la extensión de una serie finita de experimentos independientes a una sucesión infinita, uno puede intuir que los experimentos de la sucesión infinita serán también independientes. En correspondencia con el modelo de Bernoulli esto significa, que si denotamos ahora por A_i (antes \tilde{A}_i) la ocurrencia de éxito en el experimento i , entonces los eventos de la sucesión $\{A_i\}$ son independientes. Además, $P(A_i) = p$, $i \geq 1$.

Note ahora que $\{(a_1, a_2, \dots)\} \subset D_n$, para todo $n \geq 1$, de donde

$$P((a_1, a_2, \dots)) \leq P(D_n) = p^{\sum_{i=1}^n a_i} q^{n - \sum_{i=1}^n a_i} \rightarrow 0.$$

cuando $n \rightarrow \infty$. Es decir, la probabilidad de cualquier punto muestral correspondiente a una sucesión infinita de experimentos independientes de Bernoulli es igual a 0; en particular, la probabilidad de que siempre ocurra éxito es igual a 0, resultado que para $p = 1/2$ se obtuvo por diferentes vías en los Ejemplos 5.7, 5.9 y 5.11.

Ejemplo 7.6 Los eventos de la sucesión $\{A_n\}$, definidos anteriormente, son independientes, y la serie $\sum_{i=1}^{\infty} P(A_i)$ es divergente, luego por la Parte II del lema de Borel-Cantelli se halla que $P(A_n \text{ i.f.}) = 1$. Es decir, en el modelo de sucesión infinita de experimentos de Bernoulli independientes, la probabilidad de que el éxito ocurra infinitamente frecuente es igual a 1. ■

Ejemplo 7.7 Consideremos una generalización y formalización del Ejemplo 5.7. Supongamos que se realiza una sucesión infinita de experimentos independientes. Sea p , $0 < p < 1$, la probabilidad de éxito en cada uno de los experimentos.

El evento E_n , que consiste en la ocurrencia del primer éxito en el experimento número n , se representa,

$$E_n = \{(0, 0, \dots, 0, 1)\} \times \Omega_{n+1} \times \Omega_{n+2} \times \dots,$$

donde $(0, 0, \dots, 0, 1) \in \Omega^n$. Como

$$P(E_n) = P^n((0, 0, \dots, 0, 1)) = pq^{n-1},$$

entonces la probabilidad r de que el éxito ocurra por primera vez en un experimento con número de orden impar se calcula de la siguiente manera:

$$r = \sum_{n=1}^{\infty} P(E_{2n-1}) = \sum_{n=1}^{\infty} pq^{2n-2} = \frac{p}{1-q^2} = \frac{1}{1+q} = \frac{1}{2-p}.$$

Esta probabilidad es $1/q$ veces mayor que la probabilidad de que el éxito ocurra por primera vez en un experimento con número de orden par. En el Ejemplo 5.7 se trató el caso particular en que $p = 1/2$. ■

Ejercicio 1.

Considere el modelo de una sucesión infinita de experimentos de Bernoulli independientes. Pruebe que la probabilidad de ocurrencia infinitamente frecuente de dos éxitos consecutivos es igual a 1.

Sugerencia: Utilice el lema de Borel-Cantelli.

Observación 5 Una generalización del modelo del Ejercicio 1 es la siguiente: en una sucesión infinita de experimentos independientes, en cada uno de los cuales es posible la ocurrencia de un número finito de caracteres (por ejemplo, los tres caracteres a , b y c), la probabilidad de ocurrencia infinitamente frecuente de una cadena finita dada de caracteres consecutivos es igual a 1. ▲

7.5. Método del primer paso

En ocasiones, en la solución de problemas relacionados con sucesiones infinitas de experimentos independientes puede ser útil el llamado **método del primer paso**.

La idea general del método es la siguiente. Se considera una sucesión de experimentos independientes en cada uno de los cuales es de interés la ocurrencia o no de éxito. El primer paso se relaciona con el primer experimento; se calculan las probabilidades del suceso de interés condicionadas a la ocurrencia de éxito o fracaso en el primer paso, y luego se aplica la fórmula de la probabilidad total.

Ilustremos este método mostrando otra vía de solución para el Ejemplo 7.7. Denotemos por A la ocurrencia de éxito en un experimento con número de orden impar, y por A_i la ocurrencia de éxito en el experimento i , $i \geq 1$. El éxito puede ocurrir en un experimento de orden impar en una de las dos situaciones excluyentes siguientes: ocurre el éxito en el primer experimento, o no ocurre el éxito en el primer experimento y tampoco en el segundo y ocurre en un experimento de orden impar a partir del tercero. Denotemos $r = P(A)$. Usando la fórmula de la probabilidad total se tiene que

$$r = P(A) = P(A|A_1)P(A_1) + P(A|A_1^c \cap A_2^c)P(A_1^c \cap A_2^c).$$

Naturalmente que

$$P(A_1) = p, \quad P(A_1^c \cap A_2^c) = q^2, \quad P(A|A_1) = 1.$$

Veamos cómo hallar $P(A|A_1^c \cap A_2^c)$. Dado que el éxito no ocurre en el primer experimento y tampoco en el segundo, para que ocurra en un experimento numerado impar, debe ocurrir en un experimento impar a partir del tercero. Pero note que, como los experimentos son independientes, la situación a partir del tercer lanzamiento no depende de lo ocurrido anteriormente, y que desde el punto de vista probabilístico es idéntica a la situación inicial cuando se realiza el primer experimento. Luego en la situación descrita, la probabilidad de que ocurra el éxito en un experimento de orden impar a partir del tercero, dado que no ocurre el éxito ni en el primero ni en el segundo experimento, es también igual a r . Por consiguiente, $P(A|A_1^c \cap A_2^c) = r$, y entonces

$$r = P(A) = 1 \cdot p + r \cdot q^2,$$

de donde $r = 1/(1 + q)$.

En la siguiente sección se utiliza el método del primer paso.

7.6. Probabilidad de extinción de una población

Consideremos una población de individuos que pudieran ser, por ejemplo, personas, animales o células. Supongamos que en un momento dado la población tiene un número de individuos determinado, y diremos que estos individuos pertenecen a la generación 0. Cada uno de los miembros de la generación 0, con determinada probabilidad puede desaparecer, o dar lugar a nuevos individuos que constituyen entonces la generación 1, y así sucesivamente con las siguientes generaciones.

Denotemos por A_n el evento que consiste en que la población ya se extinguió en la generación n , $n \geq 1$, es decir, el número de individuos en esta generación es 0; en otras palabras, el evento A_n ocurre si la población se extingue en la generación n o en una generación anterior. Note que $A_n \subset A_{n+1}$. Entonces la sucesión $\{P(A_n)\}_{n \geq 0}$, es creciente y acotada, y por lo tanto su límite existe. Sea $r = \lim_n P(A_n)$. ¿Cuál es el significado de r ? Para responder esta pregunta usemos el teorema de la continuidad de la probabilidad. Puesto que la sucesión $\{A_n\}$ es creciente, se tiene

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

El suceso $A = \bigcup_{n=1}^{\infty} A_n$ ocurre si, y solo si, para algún n , $n \geq 1$, ocurre A_n , es decir, cuando la población se extingue. Luego r es la probabilidad de que la población se extinga.

Para un modelo particular de crecimiento-decrecimiento de una población, usando la fórmula de la probabilidad total y la noción intuitiva de independencia, hallaremos ahora la probabilidad de extinción de la población.

Supondremos que en la generación 0 el número de individuos es 1, y que con probabilidad q , $0 < q < 1$, cada individuo de una generación cualquiera da lugar a dos individuos para la siguiente generación, y con probabilidad $p = 1 - q$ este muere, no produciendo individuos para la siguiente generación. Asumiremos que lo que ocurre con cada individuo es independiente de lo que ocurre con los otros. Este tipo de proceso de crecimiento-decrecimiento de una población se conoce con el nombre de *proceso de ramificación*, ya que cada individuo podría generar una rama de descendientes.

Hallemos la probabilidad de que la población se extinga, $r = P(A)$. Para ello utilizaremos el *método del primer paso*. La población se extingue en las siguientes dos situaciones excluyentes: cuando el individuo de la generación 0 muere (suceso A_1); o cuando este no muere (suceso A_1^c), pero las dos ramas

de subpoblaciones producidas por los dos individuos de la generación 1 se extinguen. Por la fórmula de la probabilidad total tiene que

$$r = P(A) = P(A|A_1)P(A_1) + P(A|A_1^c)P(A_1^c).$$

Note que

$$P(A_1) = p, \quad P(A_1^c) = q, \quad P(A|A_1) = 1.$$

Hallemos $P(A|A_1^c)$. Dado el suceso A_1^c , la población total se extingue si, y solo si, se extinguen las subpoblaciones de descendientes correspondientes a los dos individuos de la generación 1. Pero desde el punto de vista probabilístico, los dos procesos de ramificación que corresponden a estos individuos se comportan de manera idéntica al proceso de ramificación original, y son independientes entre sí, por lo tanto cada uno de ellos también tiene probabilidad de extinción r . Entonces $P(A|A_1^c) = r^2$. Por consiguiente,

$$r = P(A) = 1 \cdot p + r^2 \cdot q.$$

Ahora la probabilidad r se halla como solución de la ecuación de segundo grado $qr^2 - r + p = 0$, cuyas dos soluciones se expresan

$$\frac{1 \pm \sqrt{1 - 4pq}}{2q} = \frac{1 \pm |p - q|}{2q}.$$

Si $q < 1/2$, la solución con signo positivo en el numerador, p/q , no es una probabilidad y, por consiguiente, no es válida. La solución válida para cualquier valor de q es

$$r = \frac{1 - |p - q|}{2q}.$$

De esta fórmula se halla que cuando $q \leq 1/2$, entonces $r = 1$ (la población se extingue *casi seguramente*). Por otro lado, cuando $q > 1/2$, entonces $r = p/q$ y la población se extingue con esta probabilidad r , y no se extingue con probabilidad $1 - r = 1 - p/q$.

Si, por ejemplo, $q = 4/5$, entonces $r = 1/4 = 0,25$, lo cual puede interpretarse como que de 1000 ocasiones que se observe el proceso de ramificación descrito, en aproximadamente 750 de ellas la población crecerá indefinidamente sin extinguirse.

Los procesos de ramificación constituyen un tipo de *proceso estocástico*. Como se mencionó anteriormente, estos son procesos que transcurren en el tiempo influidos por factores aleatorios. En el modelo de ramificación descrito, el tiempo es interpretado como las generaciones y toma los valores

$n = 0, 1, 2, \dots$, o sea, es discreto. En otros procesos estocásticos el tiempo coincide con el tiempo real durante el cual se observa el proceso, y puede ser continuo; por ejemplo, en un sistema de servicio de clientes, el número de clientes en la cola en un instante de tiempo t , que está influido por diferentes factores aleatorios, es un proceso estocástico que transcurre en un tiempo continuo.

7.7. Problemas y ejercicios

1. Se lanzan al azar de manera independiente cinco monedas, con probabilidades de aparición de escudo, p_i , $i = 1, 2, 3, 4, 5$, respectivamente. Halle la probabilidad de que se observen al menos dos escudos.
2. Un sistema está constituido por n componentes en paralelo (el sistema falla si, y solo si, fallan todos sus componentes). Calcule el número mínimo de componentes que garantiza una probabilidad de fallo del sistema menor que $\alpha = 10^{-3}$, en un intervalo de tiempo dado, si la probabilidad de funcionamiento de cada componente en ese intervalo es $p = 0,9$. Considere los fallos de los componentes independientes entre sí. **R:** $n = 4$.
3. Se lanza una moneda al azar hasta que aparezca el escudo en dos lanzamientos consecutivos. Suponga que la probabilidad de aparición del escudo en cada uno de los lanzamientos es igual a p y que todas las pruebas (lanzamientos) son independientes. Halle la probabilidad de que el número de lanzamientos sea igual a cuatro.
4. Halle la probabilidad p_n de que en una serie de n experimentos independientes de Bernoulli, con probabilidad de éxito p , no ocurra el patrón 1 0, es decir, éxito en una prueba y fracaso en la siguiente.
5. Un sistema constituido por n componentes se llama *sistema k de n* si funciona cuando funcionan al menos k , $1 \leq k \leq n$, de sus componentes. Suponga que la probabilidad de funcionamiento de cada componente hasta el instante t es igual a p . Asuma que los fallos de los componentes son independientes entre sí. Halle el valor de p para el cual la probabilidad de funcionamiento hasta el instante t del sistema a) es mayor que la del sistema b), o la del sistema b) mayor que la del sistema a), en los siguientes casos:
 1. a) 2 de 3 b) 1 de 2.

2. a) 2 de 3 b) 1 de 1.
3. a) 3 de 4 b) 1 de 1.
6. Un sistema está compuesto por n componentes. El sistema funciona cuando al menos uno de los componentes funciona. Los fallos de los componentes son independientes. La probabilidad de funcionamiento de cada uno de los componentes hasta el instante t es igual a p_i , $i = 1, 2, \dots, n$. Si el sistema está funcionando en el instante t , ¿cuál es la probabilidad de que el componente i esté funcionando en ese instante?
7. Considere dos monedas, I y II, con probabilidades de éxito $1/4$ y $3/4$, respectivamente. Se selecciona al azar una de las monedas y se lanza tres veces, en dos de las cuales el resultado fue éxito y en la tercera fracaso. ¿Cuál es la probabilidad de que la moneda elegida haya sido la I? **R:** $1/4$.
8. Dos fábricas, I y II, producen lotes de tamaño 100 de un mismo artículo. Se ha estimado que el por ciento de artículos de calidad superior que producen es 10 y 30 por ciento, respectivamente. Se utilizó un lote, pero se desconoce de cual fábrica proviene. Para decidir cuál fábrica lo produjo se estableció la siguiente regla de decisión. Si el lote tiene 20 o más artículos de calidad superior, se decide que pertenece a la fábrica II, en caso contrario se concluye que proviene de la I. Calcule las probabilidades de los dos tipos de error en que podría incurrirse con esta regla.
9. *Se realizan n experimentos independientes en cada uno de los cuales la probabilidad de éxito es p . Halle la probabilidad r de que el número de éxitos en los n experimentos sea un número par. **R:** $r = \frac{1+(1-2p)^n}{2}$. (Note que $r < 1/2$ cuando $p > 1/2$ y n es impar).
10. Dos cajas contienen solo tres tipos de artículos, I, II y III. Una de las cajas contiene 1 de tipo I, 4 de tipo II y 5 de tipo III, y la otra caja contiene 5 de tipo I, 4 de tipo II y 1 de tipo III. Se selecciona al azar una de las cajas y de esta se seleccionan al azar y de manera independiente 6 artículos. Halle la probabilidad de que la caja seleccionada contenga los siguientes números de artículos:
- a) 2 de tipo I, 2 de tipo II y 2 de tipo III.
- b) 2 de tipo I, 3 de tipo II y 1 de tipo III.
- c) 1 de tipo I, 4 de tipo II y 1 de tipo III.

11. La probabilidad de ganar una jugada es p para cualquiera de dos jugadores, I y II. Comienza a jugar I, y si pierde, entonces juega II. Continúan pasando las jugadas hasta que uno de los dos gana, y en ese momento finaliza la partida. ¿Cuál es la probabilidad r de que I gane la partida? Utilice el método del primer paso. **R:** $r = 1/(2 - p)$.
12. Sean A y B dos sucesos excluyentes en un experimento y $P(A) + P(B) < 1$. Supongamos que se realiza una sucesión de experimentos independientes. Pruebe que la probabilidad de que A ocurra antes que B es igual a $P(A)/[P(A) + P(B)]$.
13. *(C. Huygens). Dos jugadores, I y II, lanzan alternativamente al azar dos dados equilibrados. El jugador I gana si en un lanzamiento obtiene una suma de 5 puntos antes de que el jugador II obtenga una suma de 6 puntos, en caso contrario gana el jugador II. Comienza el juego I, ¿cuál es la probabilidad de que gane el jugador I? **R:** $9/19$.
14. Considere el proceso de ramificación descrito en la Sección 7.6.
 - a) Calcule la probabilidad q_k , de que la cantidad de individuos en la generación 2 sea igual a k , $k = 0, 1, 2, \dots$
 - b) Implemente un programa de simulación en lenguaje Python para calcular aproximadamente las probabilidades q_k , $k = 0, 1, 2, \dots$. Tome $q = 4/5$.

COMPLEMENTOS

8.1. Conjunto no medible

Para todos los elementos x, y del intervalo $[0, 1]$ definamos la siguiente relación de equivalencia: $x \sim y$ si $x - y$ es un número racional. Claro que $x - y \in [-1, 1]$.

Como la diferencia entre dos elementos de clases de equivalencia diferentes es un número irracional, pues si fuera racional pertenecerían a la misma clase, y por otro lado, la diferencia entre elementos de la misma clase es racional, entonces las clases son conjuntos numerables y la colección de estas es no numerable.

Seleccionemos un único elemento de cada clase de equivalencia para formar un subconjunto numerable de $[0, 1]$ que denotaremos por A . Esto es posible por el axioma de selección. Consideremos la sucesión $\{r_n\}$, $n \geq 1$, de todos los números racionales del intervalo $[-1, 1]$ diferentes entre sí. Sea

$$A_n = r_n + A = \{r_n + x : x \in A\}, \quad n \geq 1.$$

Lema 1.

Se cumplen las siguientes dos propiedades:

1. $A_i \cap A_j = \emptyset$, si $i \neq j$.
2. $[0, 1] \subset \cup_{n=1}^{\infty} A_n \subset [-1, 2]$.

Demostración:

1. Supongamos que $x \in A_i \cap A_j$ para $i \neq j$. Entonces existen $x_i, x_j \in A$ y dos números racionales r_i y r_j de la sucesión $\{r_n\}$ tales que $x = r_i + x_i = r_j + x_j$, de donde se deduce que $x_i - x_j$ es un número racional. Por lo tanto, $x_i \sim x_j$, y esto a su vez implica que $x_i = x_j$, y entonces $r_i = r_j$. Luego $i = j$, lo cual contradice la suposición $i \neq j$.

2. Si $x \in [0, 1]$, existe $y \in A$ tal que $x \sim y$ y $r = x - y$ es un número racional igual a un elemento de la sucesión $\{r_n\}$, digamos $r = r_k$. Entonces $x = r_k + y \in A_k$ y, por consiguiente, $x \in \cup_{n=1}^{\infty} A_n$.

Sea ahora $x \in \cup_{n=1}^{\infty} A_n$. Entonces existe un índice m , $m \geq 1$, tal que $x \in A_m$ y $x = r_m + y$, con $y \in A \subset [0, 1]$ y $r_m \in [-1, 1]$. De donde, $x \in [-1, 2]$. ■

Proposición 2.

El conjunto A , definido anteriormente, es no medible con respecto a la medida de longitud usual λ .

Demostración:

Sobre los subconjuntos de $[0, 1]$ medibles con respecto a λ , esta medida cumple con las propiedades correspondientes a una probabilidad. Si suponemos que A es medible según λ , entonces de la definición de A_n se obtiene que A_n es un conjunto medible y $\lambda(A_n) = \lambda(A)$ para todo $n \geq 1$. Luego, usando el Lema 1, se tiene que

$$1 = \lambda([0, 1]) \leq \sum_{n=1}^{\infty} \lambda(A_n) \leq 3.$$

Pero esto no es posible, pues la suma

$$\sum_{n=1}^{\infty} \lambda(A_n) = \sum_{n=1}^{\infty} \lambda(A)$$

debe ser 0 o ∞ y, por lo tanto, A es no medible. ■

8.2. Sobre un espacio muestral infinito no numerable

Teorema 3.

Sean el espacio muestral Ω infinito no numerable. Entonces la probabilidad correspondiente a cada $\omega \in \Omega$ es igual a 0, excepto quizás en los puntos de un subconjunto finito o infinito numerable de Ω .

Demostración:

Supongamos que el conjunto $A = \{\omega \in \Omega : P(\omega) > 0\}$ es no numerable. Consideremos los conjuntos

$$A_n = \left\{ \omega \in A : \frac{1}{n+1} < P(\omega) \leq \frac{1}{n} \right\}, \quad n \in \mathbb{N}.$$

Alguno de estos subconjuntos de A debe ser diferente del conjunto vacío, pues $P(\omega) > 0$ para todo $\omega \in A$. Además, al menos uno de los conjuntos A_n debe ser infinito no numerable, puesto que si A_n fuera finito o infinito numerable para todo $n \in \mathbb{N}$, entonces $A = \bigcup_{n=1}^{\infty} A_n$ sería numerable, y se está suponiendo que A es no numerable. Sea A_{i_0} infinito no numerable. De este conjunto puede extraerse un subconjunto infinito numerable $A_{i_0}^*$. Entonces,

$$P(A) \geq P(A_{i_0}) \geq P(A_{i_0}^*) = \sum_{\omega \in A_{i_0}^*} P(\omega) > \sum_{\omega \in A_{i_0}^*} \frac{1}{i_0 + 1} = \infty,$$

lo cual contradice que $P(A) \leq 1$. Luego la suposición de que A es no numerable no es verdadera, y por lo tanto A es, o finito, o infinito numerable, o el conjunto vacío. ■

Note que el teorema anterior es válido si Ω se sustituye por cualquier subconjunto suyo infinito no numerable.

Un caso de espacio muestral infinito no numerable, en el cual no existen puntos muestrales cuya probabilidad correspondiente es mayor que cero, es cuando se considera la probabilidad geométrica asociada al intervalo $\Omega = [0, 1]$. Un ejemplo donde Ω es infinito no numerable y contiene un subconjunto infinito numerable A , tal que $P(\omega) > 0$ para todo $\omega \in A$, es el siguiente:

$$\Omega = (0, 1], \quad A = \{\omega_1, \omega_2, \dots\}, \quad \omega_i = 1/i, \quad P(\omega_i) = 1/2^i, \quad i = 1, 2, \dots$$

8.3. Infinitos experimentos independientes

Nos limitaremos a examinar el caso de una sucesión infinita de experimentos de Bernoulli independientes. El caso correspondiente al modelo multinomial se puede tratar de manera análoga.

Consideremos el modelo de Bernoulli $(\Omega^n, \mathcal{P}(\Omega^n), P^n)$ de una serie de n experimentos independientes, donde $\Omega^n = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$, $\Omega_i = \{0, 1\}$, $1 \leq i \leq n$, y para todo $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega^n$, se tiene

$$P^n(\omega) = p^{\sum_{i=1}^n \omega_i} q^{n - \sum_{i=1}^n \omega_i},$$

con $0 < p < 1$, $q = 1 - p$.

Extenderemos este modelo a un modelo de una sucesión infinita de experimentos independientes. Definamos sobre una clase de subconjuntos de Ω^∞ una probabilidad P , de manera tal que propiedades de P^n sobre los subconjuntos de Ω^n se extiendan como propiedades de P sobre esa clase. Sea

$$C(B^n) = B^n \times \Omega_{n+1} \times \Omega_{n+2} \times \dots, \quad B^n \subset \Omega^n, \quad n \geq 1.$$

Definamos,

$$P(C(B^n)) = P^n(B^n), \quad B^n \subset \Omega^n, \quad n \geq 1.$$

Similar al caso de la medida habitual de volumen λ , según la cual no para todo subconjunto de \mathbb{R}^n está definido el volumen, en este caso no es posible extender P a todos los subconjuntos de Ω^∞ . Como toda probabilidad debe estar definida sobre una σ -álgebra, y la colección $\{C(B_n)\}$ no lo es (en general la unión de elementos de esta colección no es un elemento de la misma), entonces P será extendida a la σ -álgebra generada por esta colección, que denotaremos por $\mathcal{F}_\infty = \sigma(\{C(B^n)\})$.

Note que

$$P^{n+1}(B^n \times \Omega^{n+1}) = P^n(B^n), \quad B^n \subset \Omega^n, \quad n \geq 1. \quad (8.1)$$

Tiene lugar el siguiente teorema, el cual no demostraremos, y formularemos solo en los términos del modelo que estamos examinando.

Teorema 4 (Teorema de extensión de Kolmogorov).

Si se cumple la propiedad (8.1), entonces existe una única medida de probabilidad P definida sobre el espacio medible $(\Omega^\infty, \mathcal{F}_\infty)$ tal que

$$P(C(B^n)) = P^n(B^n), \quad B^n \subset \Omega^n, \quad n \geq 1.$$

El espacio de probabilidad $(\Omega^\infty, \mathcal{F}_\infty, P)$ es el espacio producto correspondiente a una sucesión infinita de experimentos de Bernoulli independientes.

Examinemos a continuación tres conjuntos de interés que pertenecen a \mathcal{F}_∞ y calculemos sus probabilidades.

1. El conjunto $A_i = \{(\omega_1, \omega_2, \dots) \in \Omega^\infty : \omega_i = 1\}$, $i \geq 1$, representa la ocurrencia de éxito en el experimento i , $i \geq 1$.

Sea, para n dado, $n \geq 1$,

$$\tilde{A}_i^n = \{(\omega_1, \omega_2, \dots, \omega_n) \in \Omega^n : \omega_i = 1\}.$$

(En la Sección 7.2, para n dado, el suceso \tilde{A}_i^n fue denotado por \tilde{A}_i .)

Como $A_i = C(\tilde{A}_i^n) \in \mathcal{F}_\infty$, entonces por el teorema de extensión de Kolmogorov,

$$P(A_i) = P(C(\tilde{A}_i^n)) = P^n(\tilde{A}_i^n) = p,$$

para todo i , $1 \leq i \leq n$.

De la independencia conjunta en el modelo de Bernoulli de los sucesos $\tilde{A}_1^n, \tilde{A}_2^n, \dots, \tilde{A}_n^n$, para cada n , $n \geq 2$, no es difícil verificar la independencia de los eventos de la sucesión $\{A_i\}$.

2. Comprobemos que los conjuntos puntuales de Ω^∞ pertenecen a \mathcal{F}_∞ y que su probabilidad es igual a 0. Sea

$$(a_1, a_2, \dots) \in \Omega^\infty.$$

Tomemos

$$B^n = \{(a_1, a_2, \dots, a_n)\}, \quad n \geq 1.$$

Por el teorema de extensión,

$$P(C(B^n)) = P^n(B^n) = p^{\sum_{i=1}^n a_i} q^{n - \sum_{i=1}^n a_i}. \quad (8.2)$$

Por otro lado,

$$\{(a_1, a_2, \dots)\} = \bigcap_{n=1}^{\infty} C(B^n) \in \mathcal{F}_{\infty}.$$

Entonces, como la sucesión de eventos $\{C(B^n)\}$, $n \geq 1$, es decreciente, de la propiedad de continuidad de la probabilidad y de (8.2), obtenemos

$$P(\{(a_1, a_2, \dots)\}) = \lim_n P(C(B^n)) = \lim_n p^{\sum_{i=1}^n a_i} q^{n - \sum_{i=1}^n a_i} = 0.$$

3. Sea $B^n = \{(0, 0, \dots, 0, 1)\} \subset \Omega^n$. Entonces $E_n = C(B^n) \in \mathcal{F}_{\infty}$. El conjunto E_n representa la ocurrencia de éxito por primera vez en el experimento con número de orden n , $n \geq 1$. Por el teorema de extensión,

$$P(E_n) = P(C(B^n)) = P^n(B^n) = pq^{n-1}, \quad n \geq 1.$$

Parte II

Variables aleatorias

VARIABLE ALEATORIA DISCRETA

El término variable aleatoria responde a una tradición histórica, en realidad las variables aleatorias son funciones y no variables. Una variable aleatoria es una función definida sobre un espacio muestral Ω con valores en \mathbb{R} . Si Ω es discreto, es decir, finito o infinito numerable, no es necesario exigir ninguna otra condición, pero como veremos más adelante, la situación es diferente cuando Ω es infinito no numerable.

Desde el punto de vista práctico, las variables aleatorias caracterizan de forma cuantitativa los resultados de los experimentos aleatorios.

En este capítulo se considerará el espacio de probabilidad $(\Omega, \mathcal{P}(\Omega), P)$, donde Ω es discreto.

9.1. Variable aleatoria. Distribución de probabilidad

Definición 1.

Sea Ω discreto. Toda función $X : \Omega \rightarrow \mathbb{R}$ se denomina **variable aleatoria**.

Las variables aleatorias serán denotadas por las letras mayúsculas X, Y ,

Z , etc., sin subíndices o con subíndices.

Es obvio que con esta definición la suma y el producto de un número finito de variables aleatorias es una variable aleatoria, y también lo es el cociente de dos variables aleatorias cuando este está definido. Además, para toda función $g : \mathbb{R} \rightarrow \mathbb{R}$, la función compuesta $g(X(\omega))$ es una variable aleatoria.

Ejemplo 9.1 Consideremos una serie de n experimentos en cada uno de los cuales es de interés la ocurrencia o no (éxito o fracaso) de cierto suceso. El espacio muestral Ω correspondiente puede representarse,

$$\Omega = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{0, 1\}, i = 1, \dots, n\}.$$

La variable aleatoria $X_i : \Omega \rightarrow \mathbb{R}$, definida como

$$X_i(\omega) = \omega_i, \quad \omega \in \Omega,$$

toma los valores 1 o 0 si en el experimento i , $1 \leq i \leq n$, ocurre éxito o fracaso, respectivamente. X_i toma el valor 1 en 2^{n-1} puntos muestrales y el valor 0 en los restantes 2^{n-1} puntos.

Definamos ahora

$$X(\omega) = \sum_{i=1}^n \omega_i, \quad \omega \in \Omega.$$

Esta función es una variable aleatoria que representa el número de éxitos en los n experimentos, y toma el valor k en $\binom{n}{k}$ puntos de Ω , $k = 0, 1, \dots, n$. Note que

$$X = \sum_{i=1}^n X_i.$$

■

De una variable aleatoria interesan no solo sus valores, sino también las probabilidades con que son tomados estos valores.

Sea $B \subset \mathbb{R}$. Denotemos por $X^{-1}(B)$ la preimagen de B mediante la variable aleatoria X , es decir,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}.$$

La probabilidad $P(\{\omega \in \Omega : X(\omega) \in B\})$, de que X tome valores en el conjunto B , será denotada abreviadamente como $P(X \in B)$. En particular, la probabilidad de que X tome el valor a ,

$$P(\{\omega \in \Omega : X(\omega) = a\}) = P(X^{-1}\{a\}),$$

se denotará por $P(X = a)$, y la probabilidad,

$$P(\{\omega \in \Omega : X(\omega) > a\}) = P(X^{-1}((a, +\infty)))$$

por $P(X > a)$. De manera semejante se denotarán otros casos similares.

Sean a_1, a_2, \dots los valores de una variable aleatoria X y denotemos $p_k = P(X = a_k), k \geq 1$.

Definición 2.

A la función $f(x)$, definida sobre los valores a_1, a_2, \dots de una variable aleatoria X tal que $p_k = P(X = a_k) = f(a_k)$, que establece la correspondencia entre cada valor a_k de X y la probabilidad p_k de que X tome este valor, se le denomina **función de probabilidad o distribución de probabilidad** de X .

La distribución de probabilidad de una variable aleatoria X se puede representar también por la sucesión $\{(a_k, p_k)\}, k \geq 1$.

Puesto que $X^{-1}(\{a_k\}), k \geq 1$, es una colección de eventos disjuntos, y

$$\bigcup_{k=1}^{\infty} X^{-1}(\{a_k\}) = \Omega,$$

entonces de los axiomas de la probabilidad se obtiene la siguiente proposición.

Proposición 1.

Toda función de probabilidad $f(x)$ cumple:

$$f(a_k) \geq 0, k \geq 1, \quad \text{y} \quad \sum_{k=1}^{\infty} f(a_k) = 1.$$

Como una variable aleatoria se caracteriza no solo por ser una función de Ω en \mathbb{R} , sino además por su distribución de probabilidad, entonces a veces se dice que la variable aleatoria está definida sobre el espacio de probabilidad $(\Omega, \mathcal{P}(\Omega), P)$.

Definición 3.

Se dice que una variable aleatoria X tiene **distribución de Bernoulli** (o que es una **variable de Bernoulli**) con parámetro p , y se denota $X \sim \text{Bernoulli}(p)$, si X toma solo los valores 1 y 0 con probabilidades respectivas, p y $q = 1 - p$, $0 < p < 1$.

La función de probabilidad correspondiente a una variable aleatoria de Bernoulli se expresa,

$$f(k) = p^k q^{1-k}, \quad k = 0, 1.$$

Note que $\sum_{k=0}^1 f(k) = 1$.

Definición 4.

Se dice que una variable aleatoria X tiene **distribución binomial** con parámetros n y p , y se denota $X \sim \text{Binomial}(n, p)$, si X representa el número de éxitos en el modelo de Bernoulli de n experimentos independientes en cada uno de los cuales la probabilidad de éxito es igual a p , $0 < p < 1$.

La función de probabilidad $f(k)$ de una variable aleatoria $X \sim \text{Binomial}(n, p)$, se expresa,

$$f(k) = P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad q = 1 - p, \quad k = 0, 1, \dots, n.$$

Se cumple que

$$\sum_{k=0}^n f(k) = (p + q)^n = 1.$$

En el caso particular $X \sim \text{Binomial}(1, p)$, la variable aleatoria X es una variable de Bernoulli.

Si en el Ejemplo 9.1 la probabilidad de éxito en cada uno de los experimentos es igual a p , entonces $X_i \sim \text{Bernoulli}(p)$, $i = 1, 2, \dots, n$; y si los experimentos se asumen independientes, entonces la variable aleatoria X tiene

distribución Binomial(n, p). Calculemos, como ejemplo, la probabilidad de que X tome un valor menor que 2:

$$P(X < 2) = P(X = 0) + P(X = 1) = q^n + npq^{n-1}.$$

En las aplicaciones a veces se desconoce el espacio muestral Ω , o cómo está definida la variable aleatoria, y se trabaja entonces solo con sus valores y correspondientes probabilidades; en otras ocasiones incluso no es importante conocer cómo está definida la variable aleatoria. En este sentido el siguiente teorema tiene gran importancia.

Teorema 2.

Dada una sucesión de pares $\{(a_i, p_i)\}$, tal que $a_i \in \mathbb{R}$, $p_i \geq 0$, $i \geq 1$, y $\sum_{i=1}^{\infty} p_i = 1$, existen un espacio de probabilidad (Ω, \mathcal{F}, P) , y una variable aleatoria X definida sobre este, con distribución de probabilidad, $p_i = f(a_i) = P(X = a_i)$, $i \geq 1$.

Demostración:

Tomemos $\Omega = \{a_1, a_2, \dots\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, y $X(a_i) = a_i$, $i \geq 1$, y definamos la probabilidad P sobre Ω como $P(a_i) = p_i$, $i \geq 1$. ■

El Teorema 2 muestra de que manera, si el conjunto de valores de una variable aleatoria es finito o infinito numerable, incluso cuando la variable esté definida sobre un espacio muestral infinito no numerable, esta puede redefinirse sobre un espacio muestral discreto.

Una variable aleatoria cuyo conjunto de valores es finito o infinito numerable se denomina **variable aleatoria discreta**.

Ilustremos el Teorema 2 con el siguiente ejemplo. El espacio muestral correspondiente al modelo de una sucesión infinita de experimentos con dos resultados posibles en cada experimento, éxito o fracaso, es un conjunto infinito no numerable. En el Ejemplo 7.7 se obtuvo que, dada una sucesión infinita de experimentos independientes en cada uno de los cuales la probabilidad de éxito es igual a p , la probabilidad p_k de que el primer éxito ocurra en el experimento con número de orden k es igual a $p_k = pq^{k-1}$, $k \geq 1$, $q = 1 - p$. Se tiene entonces una sucesión de pares $\{(k, p_k)\}$ con $\sum_{k=1}^{\infty} p_k = 1$. Tomemos $\Omega = \{1, 2, \dots\}$, $P(k) = p_k$ y $X(k) = k$, $k \in \Omega$; entonces X representa el número de experimentos hasta el primer éxito y la función de probabilidad de X

es

$$f(k) = pq^{k-1}, \quad k \geq 1.$$

Definición 5.

Se dice que una variable aleatoria X tiene **distribución geométrica** con parámetro p , y se denota $X \sim \text{Geo}(p)$, si X representa el número de experimentos hasta el primer éxito, en una sucesión infinita de experimentos independientes con probabilidad de éxito en cada uno de ellos igual a p , $0 < p < 1$.

El nombre de distribución geométrica proviene de la forma de la función de probabilidad correspondiente, la cual es una progresión geométrica.

Como un ejemplo de cálculo con la distribución geométrica, hallemos la probabilidad de que el número de experimentos hasta la ocurrencia del primer éxito sea mayor que j :

$$P(X > j) = \sum_{k=j+1}^{\infty} f(k) = \sum_{k=j+1}^{\infty} pq^{k-1} = q^j, \quad j \geq 1.$$

Esta probabilidad también puede calcularse notando que el primer éxito ocurre después del experimento número j si, y solo si, en cada uno de los primeros j experimentos no ocurre el éxito, y que este último suceso tiene probabilidad q^j .

En el siguiente ejercicio se propone demostrar un resultado intuitivamente obvio. Dado que en una sucesión infinita de experimentos independientes el primer éxito no ocurre en los primeros n experimentos, entonces se “olvidan” estos n experimentos, de manera tal que el número de experimentos contados desde el experimento $n+1$ hasta el primer éxito, tiene igual distribución que el número de experimentos contados desde el primer experimento hasta el primer éxito.

Ejercicio 1.

Ausencia de memoria. Sea X una variable aleatoria con distribución geométrica. Pruebe que

$$P(X = n + k | X > n) = P(X = k), \quad k = 1, 2, \dots$$

para todo $n \geq 1$.

Una generalización de la distribución geométrica es la distribución que se define a continuación.

Definición 6.

Se dice que una variable aleatoria X tiene **distribución binomial negativa** con parámetros p y m , y se denota $X \sim \text{BinomialN}(p, m)$, si X representa el número de experimentos hasta que por primera vez ocurren m éxitos en una sucesión infinita de experimentos independientes, con probabilidad de éxito en cada uno de ellos igual a p .

Si se toma $m = 1$ en la distribución binomial negativa, entonces se obtiene la distribución geométrica.

Una interpretación práctica de la distribución binomial negativa es la siguiente. Supongamos que cada objeto producido por una fábrica es defectuoso con probabilidad p , independientemente del estado de cualquier otro objeto producido. Entonces el número de objetos producidos hasta que por primera vez se acumulan m defectuosos, es una variable aleatoria con distribución binomial negativa.

Ejercicio 2.

Sea $X \sim \text{BinomialN}(p, m)$. Pruebe que la función de probabilidad de X se expresa

$$f(k) = \binom{k-1}{m-1} p^m q^{k-m}, \quad q = 1-p, \quad k = m, m+1, \dots$$

Si $X \sim \text{BinomialN}(p, m)$, entonces la variable aleatoria $Y = X - m$ representa el número de fracasos antes de obtener m éxitos. A la distribución de la variable Y también se le llama distribución binomial negativa. En este texto, cuando no se aclare a cuál distribución binomial negativa se hace referencia, se entenderá que es la correspondiente a la variable aleatoria X . La función de probabilidad de Y es

$$f(k) = P(Y = k) = P(X = k + m) = \binom{k+m-1}{m-1} p^m q^k, \quad k = 0, 1, \dots$$

Para el coeficiente binomial de la expresión anterior se tiene,

$$\begin{aligned} \binom{k+m-1}{m-1} &= \frac{(k+m-1)!}{k!(m-1)!} = \frac{(k+m-1)(k+m-2) \cdots m}{k!} \\ &= (-1)^k \frac{-m(-m-1)(-m-2) \cdots (-m-k+1)}{k!} \\ &= (-1)^k \binom{-m}{k}, \quad k = 1, 2, \dots \end{aligned}$$

Note ahora que

$$(1-q)^{-m} = \sum_{k=0}^{\infty} (-1)^k \binom{-m}{k} q^k. \quad (9.1)$$

Entonces

$$\sum_{k=0}^{\infty} f(k) = \sum_{k=0}^{\infty} (-1)^k \binom{-m}{k} p^m q^k = p^m (1-q)^{-m} = 1.$$

El nombre de distribución binomial *negativa* proviene de la fórmula 9.1.

Definición 7.

Se dice que una variable aleatoria X tiene **distribución de Poisson** con parámetro a , $a > 0$, y se denota $X \sim \text{Poisson}(a)$, si X tiene función de probabilidad

$$f(k) = P(X = k) = \frac{e^{-a} a^k}{k!}, \quad k = 0, 1, 2, \dots$$

No es difícil comprobar que

$$\sum_{k=0}^{\infty} f(k) = \sum_{k=0}^{\infty} \frac{e^{-a} a^k}{k!} = 1.$$

El siguiente ejercicio muestra que en el modelo de Bernoulli, cuando el número de experimentos n es grande y la probabilidad de éxito p es pequeña, la distribución del número total de éxitos, $\text{Binomial}(n, p)$, se aproxima a la distribución de Poisson con parámetro np .

Ejercicio 3.

Pruebe que si $np_n \rightarrow a > 0$ cuando $n \rightarrow \infty$, entonces

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \rightarrow \frac{e^{-a} a^k}{k!}, \quad k = 0, 1, \dots,$$

cuando $n \rightarrow \infty$.

Cuando n es grande y p es pequeño, del ejercicio anterior se obtiene la siguiente fórmula aproximada para la distribución binomial

$$\binom{n}{k} p^k (1 - p)^{n-k} \approx \frac{e^{-np} (np)^k}{k!}, \quad k = 0, 1, \dots, n.$$

El resultado del Ejercicio 3 explica que, con frecuencia en las aplicaciones, la distribución de Poisson se manifiesta cuando una variable aleatoria representa el número de veces que ocurre un suceso dado, si existe un gran

número n de fuentes emisoras de este suceso las cuales son relativamente independientes entre sí, y la probabilidad p de que cada fuente genere el suceso es pequeña. Son numerosos los problemas prácticos con estas características, citemos solo algunos ejemplos: el número de accidentes de aviación mundiales en un periodo de tiempo dado; el número de accidentes en una carretera durante determinado mes; el número de errores tipográficos en la escritura de un documento.

Definición 8.

Se dice que una variable aleatoria X tiene **distribución hipergeométrica** con parámetros N, M, n , y se denota $X \sim \text{Hipergeo}(N, M, n)$, si X tiene función de probabilidad

$$f(k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad \max(0, n - (N - M)) \leq k \leq \min(n, M),$$

donde además, N, M y n son numeros naturales y, por supuesto, $\max\{n, M\} \leq N$.

Supongamos que de un lote con N artículos, de los cuales M son defectuosos, se selecciona al azar, y sin reposición, una muestra de tamaño n , entonces la variable aleatoria X , que representa el número de artículos defectuosos de la muestra, tiene distribución hipergeométrica, o sea, $X \sim \text{Hipergeo}(N, M, n)$. Note que si la selección de artículos se realiza con reposición, entonces el número de artículos defectuosos en la muestra, Y , tiene distribución binomial, $Y \sim \text{Binomial}(n, M/N)$ (vea los Ejemplos 3.8 y 3.9).

El nombre “Hipergeométrica” tiene una historia fuera de la teoría de las probabilidades, y está relacionado con lo siguiente. El cociente de dos términos consecutivos de la función de probabilidad de la distribución hipergeométrica no es constante, como en una progresión geométrica, sino una función de su argumento k . Para este tipo de progresión se añadió el prefijo “Hiper”.

El siguiente fragmento de código Python muestra cómo hallar valores de las funciones de probabilidad de las distribuciones binomial y de Poisson correspondientes a un argumento k . El parámetro de la distribución de Poisson se denotó por a . De forma similar se pueden hallar los valores correspondientes a las funciones de probabilidad de otras distribuciones discretas.

```
from scipy.stats import binom, poisson
x=binom.pmf(k,n,p)
y=poisson.pmf(k,a)
```

Aquí pmf significa *probability mass function*, término cuyo significado es el mismo que función de probabilidad.

9.2. Valor esperado

Definición 9.

Sea X una variable aleatoria no negativa definida sobre un espacio muestral discreto Ω . El **valor esperado** (o **esperanza matemática**) de X , denotado por EX (o $E(X)$), se define

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega).$$

De esta definición se deduce que el valor esperado de una variable no negativa siempre existe, sea finito o infinito. Además, que si X_1, X_2, \dots, X_n son variables aleatorias no negativas definidas en el mismo espacio muestral discreto, entonces

$$E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n.$$

El siguiente teorema muestra cómo calcular el valor esperado de una variable aleatoria conocida su función de probabilidad.

Teorema 3.

Sea $p_i = f(a_i)$, $i \geq 1$, la función de probabilidad de una variable aleatoria no negativa X definida en un espacio muestral discreto Ω . Entonces

$$EX = \sum_{i=1}^{\infty} a_i p_i.$$

Demostración:

Agrupando convenientemente los términos de la serie se tiene,

$$\begin{aligned} EX &= \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{i=1}^{\infty} \sum_{\omega: X(\omega)=a_i} X(\omega)P(\omega) \\ &= \sum_{i=1}^{\infty} a_i \sum_{\omega: X(\omega)=a_i} P(\omega) = \sum_{i=1}^{\infty} a_i p_i. \end{aligned}$$

■

A veces, en lugar de referirse al valor esperado de una variable aleatoria, se hace referencia al valor esperado de la distribución de la variable.

Ejemplo 9.2 A continuación se calculan los valores esperados de varias distribuciones:

1. Si $X \sim \text{Bernoulli}(p)$, entonces $EX = p$.
2. Si $X \sim \text{Binomial}(n, p)$, entonces $EX = np$.
3. Si $X \sim \text{Geo}(p)$, entonces $EX = \frac{1}{p}$.
4. Si $X \sim \text{BinomialN}(p, m)$, entonces $EX = \frac{m}{p}$.
5. Si $X \sim \text{Poisson}(a)$, entonces $EX = a$.
6. Si $X \sim \text{Hipergeo}(N, M, n)$, entonces $EX = n \frac{M}{N}$.

En efecto:

1. $EX = 0 \cdot (1 - p) + 1 \cdot p = p$.
2. X se puede representar como $X = \sum_{i=1}^n X_i$, donde $X_i \sim \text{Bernoulli}(p)$, $1 \leq i \leq n$. Luego, $EX = \sum_{i=1}^n EX_i = np$.
3. $EX = \sum_{k=1}^{\infty} k p q^{k-1} = 1/p$.
4. X se puede representar como $X = \sum_{i=1}^m X_i$, donde $X_i \sim \text{Geo}(p)$, $1 \leq i \leq m$. Luego, $EX = \frac{m}{p}$.
5. $EX = \sum_{k=0}^{\infty} k \frac{e^{-a} a^k}{k!} = a \sum_{k=1}^{\infty} \frac{e^{-a} a^{k-1}}{(k-1)!} = a$.

6. X se puede representar como $X = \sum_{i=1}^n X_i$, donde $X_i \sim \text{Bernoulli}(\frac{M}{N})$, $1 \leq i \leq n$. Luego, $EX = \sum_{i=1}^n EX_i = n\frac{M}{N}$.

■

Supongamos ahora que X es una variable aleatoria arbitraria definida en un espacio muestral discreto. Definamos las variables aleatorias no negativas

$$X^+ = \max(X, 0), \quad X^- = -\min(X, 0).$$

Se tiene que

$$X = X^+ - X^- \quad \text{y} \quad |X| = X^+ + X^-.$$

Note que los valores esperados de X^+ y X^- siempre existen, puesto que estas son variables aleatorias no negativas.

Definición 10.

Sea X una variable aleatoria definida en un espacio muestral discreto. Si $EX^+ < \infty$ o $EX^- < \infty$, el **valor esperado** de X se define

$$EX = EX^+ - EX^-.$$

Cuando $EX^+ < \infty$ y $EX^- < \infty$, se dice que X es **sumable** (o **integrable**). Si uno de los valores, EX^+ o EX^- , es finito y el otro infinito, se dice que el valor esperado es infinito.

En el caso $EX^+ = EX^- = \infty$, el valor esperado de X no existe.

Puesto que la descomposición $X = X^+ - X^-$ es única, entonces si el valor esperado de X existe, este es único.

Cuando $|X|$ es sumable, o sea $E|X| < \infty$, se dice que X es **absolutamente sumable**. Como

$$E|X| = E(X^+ + X^-) = EX^+ + EX^-,$$

entonces X es sumable si, y solo si, es absolutamente sumable.

Por otro lado, si $E|X| = \infty$, el valor esperado de $|X|$ existe, pero el valor esperado de X podría o no existir; este no existe si $EX^+ = EX^- = \infty$.

De las Definiciones 9 y 10 de valor esperado, inmediatamente se deduce la siguiente proposición.

Proposición 4.

Sea X es una variable aleatoria definida en un espacio muestral discreto Ω y $E|X| < \infty$. Entonces

$$EX = \sum_{\omega \in \Omega} X(\omega)P(\omega).$$

Cuando la serie $\sum_{\omega \in \Omega} X(\omega)P(\omega)$ es absolutamente convergente, la igualdad que establece la proposición anterior constituye una definición equivalente del valor esperado de X .

El siguiente teorema generaliza el Teorema 3 al caso de una variable aleatoria arbitraria definida en un espacio muestral discreto.

Teorema 5.

Sea $p_i = f(a_i)$, $i \geq 1$, la función de probabilidad de una variable aleatoria X definida en un espacio muestral discreto Ω , y sea g una función definida sobre el conjunto de los números reales con valores también reales. Se cumple:

a) Si g toma valores no negativos, entonces

$$Eg(X) = \sum_{i=1}^{\infty} g(a_i)p_i.$$

b) $g(X)$ es sumable si, y solo si, la serie

$$\sum_{i=1}^{\infty} g(a_i)p_i$$

es absolutamente convergente. En el caso de convergencia absoluta,

$$Eg(X) = \sum_{i=1}^{\infty} g(a_i)p_i.$$

Demostración:

a) De forma análoga a la demostración del Teorema 5 se tiene,

$$\begin{aligned} Eg(X) &= \sum_{\omega \in \Omega} g(X(\omega))P(\omega) = \sum_{i=1}^{\infty} \left(\sum_{\omega: X(\omega)=a_i} g(X(\omega))P(\omega) \right) \\ &= \sum_{i=1}^{\infty} \left(g(a_i) \sum_{\omega: X(\omega)=a_i} P(\omega) \right) = \sum_{i=1}^{\infty} g(a_i)p_i. \end{aligned}$$

b) Las igualdades en la demostración de la parte a) se conservan escribiendo $|g(X(\omega))|$ en lugar de $g(X(\omega))$, y $|g(a_i)|$ en lugar de $g(a_i)$, luego la convergencia absoluta de las series justifica el cambio de orden en las sumas, y por consiguiente se cumple el resultado enunciado. ■

Observe que en el teorema anterior el valor esperado se halla conociendo solo la distribución de probabilidad $\{(a_i, p_i)\}$, $i \geq 1$, de la variable aleatoria X , sin conocer como está definida la variable.

Del Teorema 5 se concluye, en particular, que si X es una variable aleatoria no negativa o sumable, con función de probabilidad $p_i = f(a_i)$, $i \geq 1$, entonces

$$EX = \sum_{i=1}^{\infty} a_i f(a_i) = \sum_{i=1}^{\infty} a_i p_i.$$

Luego el valor esperado de una variable aleatoria es un valor promedio, o medio, de los valores posibles de la variable aleatoria, calculado de manera ponderada de acuerdo a su distribución de probabilidad; por eso en ocasiones se hace referencia al valor esperado como **valor medio** (lo cual sería más apropiado que *valor esperado*).

Observación 6 La media aritmética usual se puede interpretar como un promedio ponderado. Veamos un ejemplo: consideremos el promedio de las notas de un grupo de n estudiantes en una asignatura. Si m_i estudiantes obtuvieron la calificación i , $i = 2, 3, 4, 5$, entonces el promedio de las notas es el promedio ponderado $\sum_{i=2}^5 i \frac{m_i}{n}$, donde a cada valor i le corresponde un “peso” igual a la frecuencia relativa m_i/n de ocurrencia de la nota i , $i = 2, 3, 4, 5$.

▲

Ejemplo 9.3 Sean $\Omega = \mathbb{N}$ y X una variable aleatoria tal que

$$X(i) = a_i, \quad p_i = f(a_i) = P(X = a_i) = \frac{1}{i(1+i)}, \quad i \in \mathbb{N}.$$

Como es conocido, $\sum_{i=1}^{\infty} p_i = 1$.

- Si $a_i = i, i \in \mathbb{N}$, entonces obviamente la serie

$$\sum_{i=1}^{\infty} i \frac{1}{i(i+1)}$$

no es absolutamente convergente. Pero el valor esperado de X existe:

$$EX = \sum_{i=1}^{\infty} i \frac{1}{i(i+1)} = \infty.$$

- Sea $a_i = i(-1)^i, i \in \mathbb{N}$. La serie

$$\sum_{i=1}^{\infty} i(-1)^i \frac{1}{i(i+1)}$$

no converge absolutamente, luego el valor esperado de X podría ser infinito, o no existir. Pero la serie converge condicionalmente y, por consiguiente, $EX^+ = EX^- = \infty$, lo cual significa que el valor esperado de X no existe.

- Si $a_i = (-1)^i, i \in \mathbb{N}$, entonces la serie

$$\sum_{i=1}^{\infty} (-1)^i \frac{1}{i(i+1)}$$

es absolutamente convergente, luego EX existe, es finito y

$$EX = \sum_{i=1}^{\infty} (-1)^i \frac{1}{i(i+1)}.$$

- Si ahora $a_{2i} = 1$ y $a_{2i-1} = -(2i-1), i \in \mathbb{N}$, entonces $EX^+ < \infty$ y $EX^- = +\infty$, luego $E|X| = \infty$ y $EX = -\infty$.

■

Consideremos dos variables aleatorias, X y Y , definidas sobre un espacio muestral discreto Ω , y a una constante real. Obviamente se cumplen las siguientes dos propiedades:

1. Si $X(\omega) = a$, para todo $\omega \in \Omega$, entonces $EX = a$.
2. Sean X y Y no negativas o sumables. Si $X(\omega) \leq Y(\omega)$, para todo $\omega \in \Omega$, entonces $EX \leq EY$.

Teorema 6.

Sean X y Y variables aleatorias definidas sobre un espacio muestral discreto, y a y b constantes reales. Si $E|X| < \infty$ y $E|Y| < \infty$, entonces $E|aX + bY| < \infty$, y

$$E(aX + bY) = aEX + bEY.$$

Demostración:

Note que $E|aX + bY| \leq |a|E|X| + |b|E|Y| < \infty$. Por otra parte,

$$\begin{aligned} E(aX + bY) &= \sum_{\omega \in \Omega} (aX + bY)(\omega)P(\omega) = \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega))P(\omega) \\ &= a \sum_{\omega \in \Omega} X(\omega)P(\omega) + b \sum_{\omega \in \Omega} Y(\omega)P(\omega) = aEX + bEY. \end{aligned}$$

■

El teorema anterior se extiende de forma natural para un número finito arbitrario de variables aleatorias.

9.3. Sobre estadística

Aunque la estimación de los parámetros de las distribuciones de probabilidad es objeto de estudio de la Estadística matemática, expliquemos brevemente algunas ideas sobre el cálculo aproximado de algunos parámetros.

Consideremos el modelo de Bernoulli de una serie de n experimentos independientes, en cada uno de los cuales la probabilidad de éxito es igual a p .

Sea la variable aleatoria X_i que toma el valor 1 si ocurre éxito en el experimento i y el valor 0 en caso contrario, $1 \leq i \leq n$. Entonces la media aritmética

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

es una variable aleatoria que representa la frecuencia relativa de ocurrencia de éxito en los n experimentos. Como $EX_i = p$, $1 \leq i \leq n$, se tiene

$$E\bar{X} = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n EX_i = p.$$

Esta propiedad, conjuntamente con la observación empírica de que la frecuencia relativa de ocurrencia de un suceso tiende a aproximarse a su probabilidad de ocurrencia p , cuando el número de experimentos n crece, son dos de las razones que fundamentan por qué tomar un valor de la frecuencia relativa \bar{X} como valor aproximado de p , cuando n es grande.

Veamos ahora una generalización de lo anterior. Se llama **muestra de tamaño** n de una variable aleatoria X a una serie de n variables aleatorias X_1, X_2, \dots, X_n , que constituyen valores observados de X y tienen igual distribución que X , luego $EX_i = EX$, $1 \leq i \leq n$. Consideremos la variable aleatoria media aritmética $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Entonces $E\bar{X} = EX$. De nuevo, esta propiedad, conjuntamente con la aproximación de los valores de \bar{X} a EX , cuando n es grande, contribuye a fundamentar que como valor aproximado de EX se tome un valor de la media aritmética \bar{X} ; el siguiente ejemplo ilustra de manera intuitiva esta aproximación.

Supongamos que X es una variable aleatoria que representa las notas en una asignatura. Sea $p_i = P(X = i)$, $i = 2, 3, 4, 5$, la distribución de probabilidad de X . Se seleccionan al azar n estudiantes, y con ellos se obtiene una muestra de tamaño n de X , es decir, variables aleatorias X_i , cuyos posibles valores corresponden a la nota del estudiante i , $1 \leq i \leq n$. Supongamos que del total de n estudiantes, m_i obtuvieron la calificación i , $i = 2, 3, 4, 5$. Entonces el valor \bar{x} de \bar{X} para esta muestra se puede expresar como $\bar{x} = \sum_{i=2}^5 i \frac{m_i}{n}$. Cuando n es grande, $\frac{m_i}{n} \approx p_i$, $i = 2, 3, 4, 5$, de donde

$$\bar{x} = \sum_{i=2}^5 i \frac{m_i}{n} \approx \sum_{i=2}^5 i p_i = EX.$$

Ejemplo 9.4 En ocasiones se puede suponer que el número de clientes que llegan cada día a cierta unidad, para requerir un servicio en un intervalo de

tiempo dado (t_1, t_2) , es una variable aleatoria X con distribución de Poisson. Resulta que el número de clientes potenciales que pueden solicitar el servicio en ese intervalo podría ser muy grande, y la probabilidad de que cada uno de ellos en particular requiera el servicio es pequeña, luego la distribución de Poisson surge aquí como aproximación de la distribución binomial (vea el Ejercicio 3). Por otro lado, como el parámetro a de una distribución de Poisson coincide con el valor esperado de esta distribución, entonces su valor, cuando n es grande, se puede tomar como aproximadamente igual a valor de la media aritmética \bar{X} de n observaciones X_i , donde X_i representan el número de clientes que solicitan el servicio en el intervalo (t_1, t_2) en el i -ésimo día, $1 \leq i \leq n$. ■

9.4. Simulación

Veamos cómo simular valores de una variable aleatoria discreta X con distribución de probabilidad

$$p_i = f(a_i) = P(X = a_i), \quad i = 1, 2, \dots$$

Sea U un número aleatorio del intervalo $(0, 1)$, y construyamos la siguiente variable aleatoria Y (vea la Figura 9.1),

$$Y = a_1, \quad \text{si } 0 \leq U < p_1;$$

$$Y = a_i, \quad \text{si } \sum_{j=1}^{i-1} p_j \leq U < \sum_{j=1}^i p_j \quad i = 2, 3, \dots$$

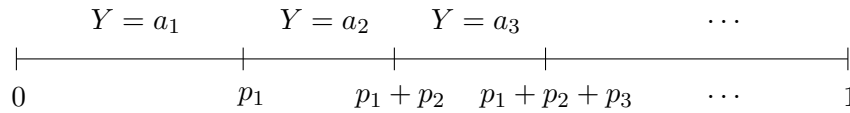


Figura 9.1: Método estándar de simulación de una variable aleatoria discreta.

La variable aleatoria Y tiene igual distribución de probabilidad que la variable aleatoria X , es decir,

$$p_i = f(a_i) = P(Y = a_i), \quad i = 1, 2, \dots$$

Entonces la generación de valores de X la podemos llevar a cabo generando valores de la variable aleatoria Y . Este método de simulación de una variable aleatoria discreta se denomina **método o algoritmo estándar**.

Ejemplo 9.5 Sea una variable aleatoria X con tres posibles valores a_1, a_2, a_3 , y distribución de probabilidad

$$P(X = a_1) = 0,1, \quad P(X = a_2) = 0,6, \quad P(X = a_3) = 0,3.$$

Para generar un valor de la variable aleatoria X se puede proceder de la siguiente manera. Generamos un valor de U , si este valor es menor que 0,1, entonces la variable aleatoria Y toma el valor a_1 y se ha generado el valor a_1 de la variable aleatoria X ; en caso contrario verificamos si el valor de U es menor que $0,1 + 0,6 = 0,7$, si esto es así, el valor generado es a_2 , y si no, el valor generado es a_3 .

Desde luego que para hacer más eficiente el algoritmo, se deben ordenar los valores de la variable aleatoria X de acuerdo con el orden de decrecimiento de las correspondientes probabilidades. En nuestro caso nos preguntáramos primero si el valor generado de U es menor que 0,6, y en caso contrario, si este valor es menor que $0,6 + 0,3 = 0,9$. ■

El número de condicionales del algoritmo estándar hasta generar un valor de una variable aleatoria X es en sí mismo una variable aleatoria Z .

Ejercicio 4.

Sea Z el número de condicionales en el algoritmo estándar hasta generar un valor de una variable aleatoria discreta X . Explique por qué si $X \sim \text{Geo}(p)$, entonces $Z \sim \text{Geo}(p)$.

Cuando el conjunto de valores de la variable aleatoria X es infinito, teóricamente las condicionales en el algoritmo estándar para generar sus valores podrían no finalizar; sin embargo, que esto suceda podría tener una probabilidad extremadamente pequeña, y en esos casos el algoritmo puede funcionar. Por ejemplo, si $X \sim \text{Geo}(p)$, con $p = 0,9$, entonces

$$q = 1 - p = 0,1 \quad \text{y} \quad P(X > 10) = q^{10} = 10^{-10},$$

por lo que en este caso es muy poco probable que con el método estándar el número de condicionales hasta generar un valor de X supere 10.

Si $X \sim \text{Geo}(p)$, entonces $EZ = 1/p$, de donde se halla que si, por ejemplo, $p = 10^{-3}$, el número medio de condicionales con el algoritmo estándar hasta generar un valor de X es $EZ = 1000$. Y si se requiere generar 1000 valores de X , entonces el número medio de condicionales es igual a 10^6 . En casos como estos es conveniente usar métodos especiales para generar los valores de X ; pero cuando en la distribución geométrica el valor de p es alto, el método estándar puede ser apropiado. Si $p = 0,9$ se tiene, $EZ \approx 1,11$.

Para la generación de valores de algunas variables aleatorias existen métodos especiales. En el siguiente ejercicio se presenta un método para la distribución geométrica.

Ejercicio 5.

Sean U un número aleatorio del intervalo $(0, 1)$, $q = 1 - p$, $0 < p < 1$, y

$$Y = \left\lceil \frac{\ln U}{\ln q} \right\rceil + 1.$$

Pruebe que $Y \sim \text{Geo}(p)$.

El siguiente fragmento de código Python muestra cómo generar un valor de las variables aleatorias con distribuciones binomial y de Poisson. El parámetro de la distribución de Poisson se denotó por a . De forma similar se pueden generar los valores correspondientes a las distribuciones de otras variables aleatorias discretas.

```
from scipy.stats import binom, poisson
x=binom.rvs(n,p)
y=poisson.rvs(a)
```

La abreviatura rvs significa de *random value sample*.

Definición 11.

Se dice que una variable aleatoria X tiene **distribución uniforme** en el conjunto $\{a_1, a_2, \dots, a_N\}$, y se denota $X \sim U\{a_1, a_2, \dots, a_N\}$, si X tiene función de probabilidad

$$f(a_k) = \frac{1}{N}, \quad k = 1, 2, \dots, N.$$

Si $X \sim U\{1, 2, \dots, N\}$, es fácil calcular que $EX = (N + 1)/2$.

Los valores de una distribución uniforme sobre el conjunto $\{a_1, \dots, a_N\}$ se pueden generar por el algoritmo estándar, pero si el valor de N es grande, es más eficiente utilizar la fórmula $\lfloor UN \rfloor + 1$ (vea el Ejercicio 3 de la Sección 2.4) para generar los índices de los valores a_i , $1 \leq i \leq N$, donde U es un número aleatorio del intervalo $(0, 1)$.

La selección de manera uniforme (aleatoria) de un subconjunto de cierto conjunto, tiene gran importancia práctica. Por ejemplo, cuando se desea valorar la eficacia de cierto medicamento, un total de N individuos se divide en dos grupos, a uno de los cuales se le aplica el medicamento, y al otro no. Al primer grupo se le llama grupo experimental, y el segundo se conoce como grupo placebo o grupo control. Si se decide que el grupo placebo sea de tamaño r , $0 < r < N$, los individuos de este grupo se seleccionan de manera que cualquiera de los $\binom{N}{r}$ grupos posibles de tamaño r tenga igual probabilidad $1/\binom{N}{r}$ de ser seleccionado. Una vez seleccionado el grupo placebo, el grupo experimental, que contiene los restantes $N - r$ individuos, quedará seleccionado también con igual probabilidad $1/\binom{N}{r}$.

La selección aleatoria del grupo de tamaño r se realiza de la siguiente manera. Primero se elige un individuo del total de N , usando la distribución uniforme; el segundo individuo se selecciona de manera uniforme entre los restantes $N - 1$; el tercero entre los restantes $N - 2$, y así sucesivamente hasta seleccionar el r -ésimo individuo. Así quedará conformado un grupo de tamaño r seleccionado con probabilidad $1/\binom{N}{r}$ de entre un total de $\binom{N}{r}$ posibles grupos.

Note que si $r > N/2$, es más eficiente seleccionar aleatoriamente el grupo experimental de tamaño $N - r$. Por otro lado, si $r = N$, el método descrito permite generar una permutación aleatoria de los números $1, 2, \dots, N$, seleccionada con igual probabilidad $1/N!$ que cualquier otra permutación de estos

números.

Ejercicio 6.

Explique por qué, en efecto, el método descrito en los párrafos anteriores, permite seleccionar de un conjunto de N elementos, un subconjunto de r elementos de entre el total de $\binom{N}{r}$ subconjuntos posibles, con igual probabilidad $1/\binom{N}{r}$. Implemente un programa en lenguaje Python para generar este subconjunto.

El siguiente ejemplo muestra cómo la distribución uniforme en un conjunto finito se podría utilizar para calcular aproximadamente sumas con un gran número N de sumandos, a_k , $1 \leq k \leq N$, los cuales podrían tener expresiones complicadas.

Ejemplo 9.6 *Método de Monte Carlo.* Veamos cómo calcular aproximadamente la suma

$$S = \sum_{k=1}^N a_k.$$

Sean $X \sim U\{a_1, a_2, \dots, a_N\}$, y X_1, X_2, \dots, X_n una muestra de X , es decir, variables aleatorias con igual distribución uniforme que X . Entonces

$$EX = \frac{1}{N} \sum_{k=1}^N a_k.$$

Por otro lado, cuando n es grande, EX es aproximadamente igual a un valor de $\frac{1}{n} \sum_{i=1}^n X_i$, o sea,

$$EX \approx \frac{1}{n} \sum_{i=1}^n x_i,$$

donde x_1, x_2, \dots, x_n son valores de X_1, X_2, \dots, X_n , respectivamente. Luego,

$$\sum_{k=1}^N a_k \approx \frac{N}{n} \sum_{i=1}^n x_i.$$

Por supuesto que este método tiene valor práctico solo si, siendo grande el valor de n , este es mucho menor que N . En algunos casos, para lograr una buena aproximación en el cálculo de la suma S , el valor de n debe ser del orden de N y entonces el método no tiene utilidad. ■

9.5. Noción sobre la teoría de juegos

Considere dos jugadores. El jugador I esconde un billete de 10 pesos o un billete de 20 pesos. El jugador II intenta adivinar cual es el billete escondido y, si acierta, obtiene ese billete; si no acierta paga 15 pesos al jugador I. Luego cada jugador tiene dos posibles estrategias: seleccionar 10 pesos o 20 pesos. La siguiente tabla expresa la matriz de “ganancias” del jugador I correspondiente a las estrategias de cada jugador. Esta matriz se denomina *matriz del juego*. Note que en algunos casos la “ganancia” es negativa.

	I_{10}	I_{20}
II_{10}	-10	15
II_{20}	15	-20

¿Cuál es la estrategia de juego más favorable al jugador I? Si se juega siempre con una estrategia fija no se puede confundir al contrario. Supongamos que el jugador I escoge la estrategia 10 pesos con probabilidad p y la estrategia 20 pesos con probabilidad $1 - p$, $0 \leq p \leq 1$. De esta manera el jugador I tiene una colección infinita de posibles estrategias, llamadas *mixtas*, determinadas por los valores de p , $0 < p < 1$. Las estrategias $p = 0$ y $p = 1$ se denominan *puras*.

Halleemos la ganancia óptima del jugador I según el criterio de optimalidad denominado *maximin*, que consiste en determinar el valor de p que hace máxima la ganancia media mínima garantizada para este jugador.

Si el jugador II escoge la estrategia 10 pesos, la ganancia del jugador I es la variable aleatoria G_{10} que toma el valor -10 con probabilidad p , y el valor 15 con probabilidad $1 - p$. El valor esperado de esta variable (ganancia media) es

$$E(G_{10}) = -10p + 15(1 - p) = 15 - 25p.$$

De manera similar, si el jugador II escoge la estrategia 20 pesos, la ganancia del jugador I es una variable aleatoria G_{20} cuyo valor esperado es

$$E(G_{20}) = 15p - 20(1 - p) = 35p - 20.$$

Si, por ejemplo, el jugador I escoge la estrategia $p = 1/5$, y el jugador II juega siempre 10, claro que I gana en promedio $E(G_{10}) = 10$; pero si II juega siempre 20, entonces I pierde en promedio 13 (gana $E(G_{20}) = -13$).

Entonces la ganancia media mínima del jugador I, o su ganancia media garantizada en un juego, para los diferentes valores de p , se halla,

$$\varphi(p) = \min(15 - 25p, 35p - 20), \quad p \in [0, 1].$$

El jugador I debe elegir un valor p_0 de p que maximiza la ganancia media garantizada,

$$\varphi(p_0) = \max_p \varphi(p).$$

Este es el llamado criterio maximin de optimización. En este caso el valor p_0 es el valor de p que satisface la ecuación

$$15 - 25p = 35p - 20,$$

de donde, $p_0 = \frac{7}{12}$. Este valor p_0 es la estrategia óptima para el jugador I. Jugando esta estrategia, cualquier otra estrategia que juegue II, el jugador I tendrá como mínimo la ganancia media $\varphi(\frac{7}{12}) = \frac{5}{12}$ pesos por jugada. Por lo tanto, la ganancia media óptima del jugador I en un juego, denominada *valor del juego*, se obtiene cuando su estrategia consiste en esconder aleatoriamente con probabilidad $\frac{7}{12}$ el billete de 10 pesos, y con la probabilidad complementaria $\frac{5}{12}$ el billete de 20 pesos. Cualquier otra estrategia que juegue I, el jugador II podría, con alguna estrategia suya, hacer que la ganancia media del jugador I sea menor, incluso que tenga pérdidas.

El valor del juego para el jugador I se interpreta de la siguiente manera. En un total de, por ejemplo, 1000 juegos, la ganancia media del jugador I, jugando con la estrategia $p_0 = \frac{7}{12}$, será aproximadamente como mínimo $1000 \frac{5}{12} = 416,7$ pesos.

Ejercicio 7.

Pruebe que la estrategia óptima para el jugador II es la misma que la del jugador I y halle el valor del juego correspondiente.

9.6. Problemas y ejercicios

1. Sea X una variable aleatoria con distribución de Poisson. Halle una fórmula para la probabilidad de que el valor de X sea un número par.

2. Investigue el crecimiento, decrecimiento y extremos de la función de probabilidad $f(k)$ de una variable aleatoria con distribución: a) binomial, b) de Poisson.
3. *Pruebe que si $\frac{M}{N} \rightarrow p$ cuando $N, M \rightarrow \infty$, entonces el límite de la función de probabilidad de la distribución Hipergeo(N, M, n), cuando $N, M \rightarrow \infty$, es igual a la función de probabilidad de la distribución Binomial(n, p).
4. Sean $X \sim \text{Binomial}(p, m)$ y $Y \sim \text{Binomial}(n, p)$. Sin realizar cálculos explique por qué $P(X > n) = P(Y < m)$.
5. Suponga que la variable aleatoria X toma los valores 0, 1, 2, y que $P(X = i) = cP(X = i - 1)$, $i = 1, 2$, para alguna constante real c . Halle EX en función solo de c .
6. Se ha estimado que el número medio de accidentes durante un período de 30 días en una carretera es 3,5. Asuma que el número de accidentes que ocurren en el transcurso de 30 días es una variable aleatoria con distribución de Poisson. Halle la probabilidad de que en ese periodo de tiempo en la carretera ocurran: a) al menos tres accidentes, b) a lo más un accidente.
7. Dos jugadores juegan una serie de partidas hasta que uno de ellos gana 2 partidas. Uno de los jugadores gana cada partida, de forma independiente, con probabilidad p , y el otro gana cada partida con probabilidad $1 - p$. Halle el valor esperado del número de partidas hasta el fin del juego, y el valor de p que lo maximiza.
8. Sea X una variable aleatoria con valores enteros no negativos. Pruebe que

$$EX = \sum_{i=1}^{\infty} P(X \geq i).$$

9. *Un dispositivo está constituido por dos componentes, 1 y 2. El fallo de cualquiera de los componentes provoca el fallo del dispositivo, pero los componentes no fallan simultáneamente; además, una vez que falla un componente el otro ya queda inactivo y no falla. Al cabo del tiempo t se observó que el dispositivo ha fallado. Se conoce que la probabilidad de fallo del componente i al cabo del tiempo t es q_i , $i = 1, 2$. Sea C_i el costo de revisión, y R_i el de reparación si esta se requiere, del componente i , $i = 1, 2$. ¿Qué relaciones deben cumplir q_i , C_i y R_i , $i = 1, 2$, para que

revisando primero el componente 1 el costo medio de reparación del dispositivo sea mínimo?

10. Sea $\{X_i\}$ una sucesión de variables aleatorias independientes igualmente distribuidas (i.i.d) con distribución $P(X_1 = 1) = p$, $P(X_1 = -1) = 1 - p$, $0 < p < 1$. Se denomina *paseo aleatorio simple* a la sucesión de variables aleatorias $\{S_n\}$, donde $S_0 = 0$ y $S_n = X_1 + \cdots + X_n$, $n \geq 1$.
 - a) Halle ES_n .
 - b) Calcule $P(S_n = k)$, $k \in \mathbb{Z}$.
 - c) Demuestre que si $p \neq 1/2$, entonces $P(S_n = 0 \text{ i.f.}) = 0$.
Sugerencia para c): Use el lema de Borel-Cantelli, Parte I.
11. *El vendedor de cierto producto compra cada unidad a 10 pesos y la vende a 15 pesos. El producto debe venderse en el día, pues en caso contrario pierde por completo sus propiedades. El vendedor sabe que en un día no vende más de 10 unidades, y ha estimado que la demanda media diaria de su producto es 5 unidades. Modelará la demanda diaria suponiendo que esta es una variable aleatoria con distribución binomial. ¿Aproximadamente cuántas unidades del producto debe comprar el vendedor para maximizar su ganancia media diaria?
12. *Estadística.* Sea $X \sim \text{Binomial}(n, p)$. La variable aleatoria X representa el número de éxitos en n experimentos independientes en cada uno de los cuales la probabilidad de éxito es p . Suponga que al realizar los n experimentos el valor observado de X es k . Una estimación razonable de p puede obtenerse hallando el valor de p que maximiza la función $L(p) = P(X = k)$ para el valor dado k (n está fijo). Esta estimación se denomina *estimación máximo verosímil*. Halle la estimación máximo verosímil de p .
13. **Estadística.* Sea $X \sim \text{Hipergeo}(N, M, n)$. La variable X podría representar el número de individuos marcados, en una muestra de tamaño n de individuos seleccionados al azar de un total de N individuos, de los cuales M están marcados. Supongamos que en la muestra se hallaron k individuos marcados. Análogamente al ejercicio anterior, halle la estimación máximo verosímil de N , es decir, el valor de N que maximiza la función $L(N) = P(X = k)$ para el valor dado k (M y n fijos).
14. Implemente en lenguaje Python un algoritmo para generar valores de una variable aleatoria con distribución $\text{Geo}(p)$, que permita además cal-

cular aproximadamente el valor de p con los datos generados. Compare esta aproximación con el valor real de p .

15. Calcule el valor exacto de la suma $\sum_{i=1}^N e^{i/N}$ para $N = 10^5$. Use el método descrito en el Ejemplo 9.6 para calcular de forma aproximada esa suma, tomando $n = 10^3$. Compare los resultados.

CAPÍTULO 10

VARIABLE ALEATORIA EN GENERAL

En el Capítulo 9 las variables aleatorias se consideraron definidas sobre un espacio muestral discreto. En este capítulo se considerarán definidas de manera general, es decir, en un espacio de probabilidad arbitrario.

10.1. Introducción

Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , para que tenga sentido referirse a la probabilidad $P(X \in B)$ de que una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$ tome valores en un subconjunto B de \mathbb{R} , es necesario que el conjunto

$$(X \in B) = X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

pertenezca a la σ -álgebra \mathcal{F} . Esto siempre sucederá si Ω es discreto, pues en este caso P está definida sobre $\mathcal{F} = \mathcal{P}(\Omega)$.

Sin embargo, si Ω es infinito no numerable, en general la probabilidad P no está definida sobre cualquier σ -álgebra en Ω . Por ejemplo, si se consideran, $\Omega = [0, 1]$, P la probabilidad geométrica, y la función $X(\omega) = \omega$, $\omega \in \Omega$, entonces cuando B es un subconjunto de Ω que no tiene longitud (vea la Sección 8.1) la probabilidad P no está definida sobre $X^{-1}(B) = B$.

Luego, tendrá validez hablar de la probabilidad de que una variable aleatoria X tome valores en un subconjunto B de \mathbb{R} , si en la definición de variable

aleatoria se exige que la preimagen de B mediante X esté contenida en la σ -álgebra \mathcal{F} sobre la cual está definida la probabilidad P . Para ello, la clase de subconjuntos de \mathbb{R} que se considerará en la definición de variable aleatoria es la llamada σ -álgebra de Borel.

10.2. Sigma-álgebra de Borel

Definición 1.

La σ -álgebra $\sigma(\{(a, b) : -\infty < a < b < \infty\})$, generada por todos los intervalos de la forma (a, b) , se denomina **σ -álgebra de Borel** o *σ -álgebra de los Borelianos* en \mathbb{R} . Esta σ -álgebra se denota $\mathcal{B}(\mathbb{R})$. Los conjuntos que pertenecen a $\mathcal{B}(\mathbb{R})$ se llaman **Borelianos**.

Puesto que

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

entonces los intervalos de la forma $[a, b]$ pertenecen a $\mathcal{B}(\mathbb{R})$. Por otra parte, los conjuntos abiertos son Borelianos, pues se pueden expresar como la unión finita o infinita numerable de intervalos abiertos. En el siguiente ejercicio se presentan algunos conjuntos Borelianos.

Ejercicio 1.

Sean $a, b \in \mathbb{R}$, $a < b$. Pruebe que el conjunto puntual $\{a\}$ y los siguientes intervalos, $(a, b]$, $[a, b)$, $(a, +\infty)$, $[a, +\infty)$, $(-\infty, b)$ y $(-\infty, b]$, pertenecen a la σ -álgebra $\mathcal{B}(\mathbb{R})$.

Como los subconjuntos puntuales de \mathbb{R} son Borelianos, es evidente que el conjunto de los números racionales es un Boreliano, por ser este la unión numerable de conjuntos puntuales, y de ahí se deduce que su complemento, el conjunto de los números irracionales, también es un Boreliano. Grosso modo, la σ -álgebra de los Borelianos se obtiene por medio de un número finito o infinito numerable de operaciones de unión, intersección y complemento de

intervalos. En realidad no es fácil dar ejemplos de subconjuntos de \mathbb{R} que no sean Borelianos.

Además de la colección de intervalos del tipo (a, b) que hemos utilizado en la definición de la σ -álgebra $\mathcal{B}(\mathbb{R})$, existen otros sistemas generadores de esta σ -álgebra. Algunos de ellos se consideran en la siguiente proposición.

Proposición 1.

Sea $a < b$. Son sistemas generadores de $\mathcal{B}(\mathbb{R})$ cada una de las colecciones de intervalos de los siguientes tipos: $[a, b]$, $[a, b)$, $(a, +\infty)$, $[a, +\infty)$, $(-\infty, b)$ y $(-\infty, b]$, y también la colección de todos los conjuntos abiertos de \mathbb{R} .

Demostración:

Solo probaremos que la colección de intervalos del tipo $[a, b]$, donde $a < b$, es un sistema generador de $\mathcal{B}(\mathbb{R})$. El resto de los casos se prueba de manera similar. Denotemos por brevedad en la escritura, $\sigma(\{[a, b]\}) = \sigma(\{[a, b] : -\infty < a < b < \infty\})$. Ya se probó que los intervalos del tipo $[a, b]$ pertenecen a $\mathcal{B}(\mathbb{R})$, luego la mínima σ -álgebra que contiene la colección de estos intervalos está contenida en $\mathcal{B}(\mathbb{R})$, es decir, $\sigma(\{[a, b]\}) \subset \mathcal{B}(\mathbb{R})$. Notemos ahora que

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

luego $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b)\}) \subset \sigma(\{[a, b]\})$. Por lo tanto, finalmente tenemos que $\sigma(\{[a, b]\}) = \mathcal{B}(\mathbb{R})$. ■

Si en general se considera \mathbb{R}^n , $n \geq 1$, se tiene la siguiente definición.

Definición 2.

Sea $-\infty < a_i < b_i < \infty$, $i = 1, 2, \dots, n$. La σ -álgebra generada por la clase de los paralelepípedos de la forma

$$\{x = (x_1, \dots, x_n) : a_i < x_i < b_i, i = 1, 2, \dots, n\},$$

se denomina σ -álgebra de los Borelianos en \mathbb{R}^n y se denota por $\mathcal{B}(\mathbb{R}^n)$.

De forma análoga a la σ -álgebra $\mathcal{B}(\mathbb{R})$, la σ -álgebra $\mathcal{B}(\mathbb{R}^n)$, $n \geq 2$, puede ser generada por la clase de los subconjuntos abiertos de \mathbb{R}^n , y por diferentes clases de paralelepípedos. También puede ser generada por la clase

$$\{B_1 \times B_2 \times \cdots \times B_n : B_i \in \mathcal{B}(\mathbb{R}), 1 \leq i \leq n\}.$$

La σ -álgebra de los Borelianos es suficientemente rica, tanto desde el punto de vista práctico, como teórico; en ella se puede definir el concepto de volumen en un sentido más general que el concepto de volumen que se obtiene a través de la integral de Riemann.

No para todos los Borelianos de $\mathcal{B}(\mathbb{R}^n)$ está definida la *medida* (longitud para $n = 1$, área para $n = 2$, etc.), en el sentido de una integral de Riemann; sin embargo, los Borelianos tienen medida (puede ser infinita) en el sentido de la llamada medida de Lebesgue.

Definición 3.

Se dice que una función μ definida sobre una σ -álgebra \mathcal{F} en Ω es una **medida**, si satisface:

1. $\mu : \mathcal{F} \rightarrow [0, \infty]$
2. $\mu(\emptyset) = 0$
3. μ es σ -aditiva.

La terna $(\Omega, \mathcal{F}, \mu)$ se denomina espacio de medida.

Observe que una medida de probabilidad P es un caso particular de medida. En general, una medida μ cumple muchas de las propiedades de una probabilidad, aunque no todas, debido a que una probabilidad toma valores solo en el intervalo $[0, 1]$. Toda medida μ cumple las propiedades de aditividad y de sub-aditividad; también, si $A, B \in \mathcal{F}$, con $B \subset A$, entonces $\mu(A \setminus B) = \mu(A) - \mu(B)$ y $\mu(B) \leq \mu(A)$.

Sea el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. En la *teoría de la medida* se demuestra que existe una medida λ , la llamada *medida de Lebesgue*, definida sobre $\mathcal{B}(\mathbb{R})$, que coincide con la longitud de los intervalos, y corresponde con la noción intuitiva de longitud. Esto es, para $-\infty < a < b < \infty$ se tiene

$$\lambda((a, b)) = \lambda([a, b]) = \lambda([a, b)) = \lambda((a, b]) = b - a.$$

La medida de Lebesgue de un conjunto puntual de $B(\mathbb{R})$ es igual a 0, luego, por la aditividad numerable, la medida de Lebesgue del conjunto de los números racionales es igual a 0. La medida de Lebesgue del intervalo $[0, 1]$ es igual a 1, y entonces también será igual a 1 la medida de Lebesgue del conjunto de los números irracionales de este intervalo, el cual es el complemento del conjunto de los números racionales del intervalo.

Observe que no es posible calcular la medida del conjunto de los números racionales del intervalo $[0, 1]$ usando la integral de Riemann, puesto que la función $g(x)$ definida como, $g(x) = 1$ si $x \in \mathbb{Q} \cap [0, 1]$, y $g(x) = 0$ en caso contrario, no es integrable según Riemann.

Anteriormente señalamos que es difícil dar ejemplos de subconjuntos de \mathbb{R} que no sean Borelianos. Pero en la Sección 8.1, se da un ejemplo de subconjunto de \mathbb{R} que no tiene medida de Lebesgue, y como todos los Borelianos son medibles según la medida de Lebesgue, esto significa que ese conjunto no es un Boreliano.

En estas notas no se estudiará el concepto general de medida, ni en particular el de medida de Lebesgue, pero esto no representará una dificultad en lo adelante, pues en los ejemplos prácticos, y en algunas cuestiones teóricas, consideraremos solo Borelianos que tienen volumen en el sentido de la integral de Riemann, que en este caso coincide con el volumen según la medida de Lebesgue.

La noción de medida de Lebesgue se generaliza de forma natural para el espacio medible $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $n \geq 2$. El modelo de la probabilidad geométrica se formaliza ahora de la siguiente manera. Sea Ω un Boreliano de \mathbb{R}^n que tiene medida de Lebesgue (volumen) finita y mayor que 0. Denotemos por $\mathcal{B}(\Omega)$ la σ -álgebra de los Borelianos contenidos en Ω ,

$$\mathcal{B}(\Omega) = \{\Omega \cap B : B \in \mathcal{B}(\mathbb{R}^n)\}.$$

La probabilidad geométrica P se define sobre $\mathcal{B}(\Omega)$ de la siguiente manera:

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)}, \quad A \in \mathcal{B}(\Omega),$$

donde $\lambda(C)$ denota la medida de Lebesgue de un conjunto $C \in \mathcal{B}(\Omega)$. El espacio de probabilidad $(\Omega, \mathcal{B}(\Omega), P)$, donde P es la probabilidad geométrica, se denomina *modelo geométrico*.

10.3. Variable aleatoria

En esta sección, para el caso de un espacio muestral arbitrario, se introducen los conceptos de variable aleatoria y de distribución de probabilidad.

Definición 4.

Sea el espacio medible (Ω, \mathcal{F}) . Una **variable aleatoria** es una función $X : \Omega \rightarrow \mathbb{R}$ tal que

$$(X \in B) = X^{-1}(B) \in \mathcal{F}, \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}).$$

A veces se dice que la variable aleatoria X está definida sobre (Ω, \mathcal{F}) , y como sobre la σ -álgebra \mathcal{F} se define una probabilidad P , entonces también se dice que la variable aleatoria está definida sobre el espacio de probabilidad (Ω, \mathcal{F}, P) .

Note, como ya fue mencionado anteriormente, que si Ω es discreto y se considera el espacio medible $(\Omega, \mathcal{P}(\Omega))$, entonces toda función real X definida sobre Ω cumple con la definición general de variable aleatoria, pues $X^{-1}(B) \in \mathcal{F} = \mathcal{P}(\Omega)$, cualquiera sea $B \subset \mathbb{R}$.

Ejercicio 2.

Sean (Ω, \mathcal{F}) un espacio medible y X una función definida sobre Ω con valores a_1, a_2, a_3, \dots . Verifique que para que X sea una variable aleatoria es necesario y suficiente que $X^{-1}(\{a_i\}) \in \mathcal{F}$, para todo $i \geq 1$.

Ejemplo 10.1 Sean el modelo geométrico, $(\Omega, \mathcal{F}, P) = ((0, 1), \mathcal{B}((0, 1)), \lambda)$, y la función $X(\omega) = 1 - \omega$, $\omega \in (0, 1)$. Es evidente que X es una variable aleatoria. La variable X podría representar el tiempo de espera, o tiempo transcurrido, desde un instante dado ω hasta el instante 1. La probabilidad de que X tome un valor en el intervalo $(1/2, 2/3)$, expresada abreviadamente como $P(1/2 < X < 2/3)$, se calcula,

$$\begin{aligned} P(1/2 < X < 2/3) &= P(\{\omega \in \Omega : 1/2 < X(\omega) < 2/3\}) \\ &= P(\{\omega \in \Omega : 1/2 < 1 - \omega < 2/3\}) \\ &= \lambda((1/2, 1/3)) = 1/6. \end{aligned}$$

■

Definición 5.

Se denomina **σ -álgebra generada o inducida** por una variable aleatoria X , y se denota por $\sigma(X)$, a la colección de eventos

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

Ejercicio 3.

Sea X una variable aleatoria definida sobre el espacio medible (Ω, \mathcal{F}) . Pruebe que, efectivamente, $\sigma(X)$ es una σ -álgebra en Ω .

En lo adelante, en ocasiones, cuando se haga referencia a una variable aleatoria, se entenderá que la variable está definida en cierto espacio de probabilidad (Ω, \mathcal{F}, P) , sin precisar este espacio.

Ejemplo 10.2 Sea c un número real. La función constante

$$X(\omega) = c, \quad \text{para todo } \omega \in \Omega,$$

es una variable aleatoria, puesto que para todo $B \subset \mathbb{R}$ se tiene

$$X^{-1}(B) = \Omega \text{ si } c \in B \text{ y } X^{-1}(B) = \emptyset \text{ si } c \notin B.$$

En este caso, la σ -álgebra inducida por X es la σ -álgebra trivial, $\sigma(X) = \{\Omega, \emptyset\}$, la cual está contenida en cualquier σ -álgebra en Ω . ■

Ejemplo 10.3 Sea $A \subset \Omega$. Consideremos la llamada **función indicadora de A** , definida como

$$I_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A. \end{cases}$$

También se usa el término *variable indicadora* en lugar de función indicadora. Note que una variable aleatoria de Bernoulli es una variable indicadora.

Si A es un evento, o sea $A \in \mathcal{F}$, entonces la función indicadora de A es una variable aleatoria. Efectivamente, para todo $B \subset \mathbb{R}$ se tiene

$$I_A^{-1}(B) = A, \text{ si } 1 \in B, 0 \notin B; \quad I_A^{-1}(B) = A^c, \text{ si } 1 \notin B, 0 \in B;$$

$$I_A^{-1}(B) = \Omega, \text{ si } 1 \in B, 0 \in B; \quad I_A^{-1}(B) = \emptyset, \text{ si } 1 \notin B, 0 \notin B.$$

La σ -álgebra inducida por la función indicadora I_A es $\sigma(I_A) = \{\Omega, \emptyset, A, A^c\}$.

Si un subconjunto A de Ω no es un evento, entonces la función indicadora I_A no es una variable aleatoria, puesto que en este caso $I_A^{-1}(\{1\}) = A \notin \mathcal{F}$. ■

10.4. Distribución de probabilidad

A continuación se introduce el concepto de distribución de probabilidad, que representa una forma general de relación entre los valores de una variable aleatoria y las probabilidades de que estos valores pertenezcan a conjuntos Borelianos.

Definición 6.

Dada una variable aleatoria X , se denomina **distribución de probabilidad** de X a la probabilidad P_X definida sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ por la fórmula

$$P_X(B) = P(X^{-1}(B)), \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}),$$

A P_X también se le llama *probabilidad inducida* por la variable aleatoria X .

A partir de un espacio de probabilidad (Ω, \mathcal{F}, P) , y de una variable aleatoria X definida sobre este espacio, se crea el nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

Ejercicio 4.

Sea X una variable aleatoria. Pruebe que P_X es una medida de probabilidad definida sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Ejemplo 10.4 Sea X una variable aleatoria tal que $X^{-1}(\{c\}) = A$ y $P(A) = 1$, donde c una constante real. (Si en particular, $A = \Omega$, entonces X es una variable aleatoria constante). La distribución de probabilidad P_X de X queda determinada como

$$P_X(B) = P(A) = 1 \text{ si } c \in B \quad \text{y} \quad P_X(B) = P(A^c) = 0 \text{ si } c \notin B,$$

cualquiera sea $B \subset \mathbb{R}$. ■

En el ejemplo anterior se considera un caso particular de variable aleatoria discreta; en general la distribución de probabilidad de una variable aleatoria discreta X queda determinada por su función de probabilidad $p_i = f(a_i)$, $i = 1, 2, \dots$. En efecto, sea $B \subset \mathbb{R}$. Entonces

$$P_X(B) = \sum_{i: a_i \in B} p_i,$$

y $P_X(B) = 0$ si $a_i \notin B$ para todo $i = 1, 2, \dots$.

Ejemplo 10.5 Consideremos el espacio medible $((a, b), \mathcal{B}((a, b)))$. Definamos la función X tal que $X(\omega) = \omega$, para todo $\omega \in (a, b)$. Es obvio que X es una variable aleatoria. Si ahora definimos una probabilidad P sobre $((a, b), \mathcal{B}((a, b)))$ de acuerdo al modelo geométrico, entonces la distribución de probabilidad P_X de X , definida sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, se determina como

$$P_X(B) = \frac{\lambda(B \cap (a, b))}{b - a}, \text{ para todo } B \in \mathcal{B}(\mathbb{R}),$$

donde λ denota la medida de Lebesgue. En particular, $P_X(B) = \lambda(B)/(b - a)$ si $B \subset (a, b)$; $P_X(B) = 1$ si $(a, b) \subset B$, y $P_X(B) = 0$ si $B \cap (a, b) = \emptyset$.

Con el modelo geométrico asumido para P , la variable aleatoria X representa un punto seleccionado al azar del intervalo (a, b) , de manera tal que la probabilidad de que X tome valores en Borelianos contenidos en el intervalo (a, b) es la misma, siempre que la medida (longitud) de estos Borelianos sea también la misma. Desde luego que la probabilidad de que X tome un valor c tal que $c \in (a, b)$, es igual a 0. ■

La distribución P_X definida en el ejemplo anterior se denomina **distribución uniforme** en el intervalo (a, b) . Para expresar que una variable aleatoria X tiene esta distribución se usa la notación $X \sim U(a, b)$.

Note que una misma distribución de probabilidad puede corresponder a diferentes variables aleatorias. Por ejemplo, si se consideran las variables aleatorias $X(\omega) = \omega$ y $Y(\omega) = 1 - \omega$, $\omega \in (0, 1)$, definidas en el espacio de probabilidad $((0, 1), \mathcal{B}((0, 1)), \lambda)$, entonces X y Y tienen igual distribución de probabilidad, uniforme en el intervalo $(0, 1)$.

10.5. Independencia de variables aleatorias

Definición 7.

Sea el espacio de probabilidad (Ω, \mathcal{F}, P) . Se dice que las variables aleatorias X_1, X_2, \dots, X_n son **variables aleatorias independientes**, si las correspondientes σ -álgebras generadas, $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$, son independientes.

Ejemplo 10.6 Consideremos el espacio de probabilidad (Ω, \mathcal{F}, P) . Los sucesos A, B y C son conjuntamente independientes si, y solo si, las σ -álgebras,

$$\{\Omega, \emptyset, A, A^c\}, \quad \{\Omega, \emptyset, B, B^c\} \quad \text{y} \quad \{\Omega, \emptyset, C, C^c\},$$

son independientes, y por lo tanto, si, y solo si, las variables aleatorias I_A, I_B y I_C son independientes.

En el modelo de Bernoulli de n experimentos independientes, las variables de Bernoulli X_1, \dots, X_n , indicadoras de la ocurrencia de éxito en cada una de los experimentos, son independientes, puesto que las correspondientes σ -álgebras $\sigma(X_1), \dots, \sigma(X_n)$ lo son. ■

Note que la definición de variables aleatorias independientes puede formularse de la siguiente forma equivalente.

Definición 8.

Sea el espacio de probabilidad (Ω, \mathcal{F}, P) . Las variables aleatorias X_1, X_2, \dots, X_n son independientes, si cualesquiera que sean los Borelianos B_1, B_2, \dots, B_n , los sucesos

$$(X_1 \in B_1), (X_2 \in B_2), \dots, (X_n \in B_n)$$

son conjuntamente independientes.

En el caso de variables aleatorias discretas, no es difícil comprobar que la definición anterior es equivalente a la independencia de las preimágenes de los conjuntos puntuales correspondientes a los valores de las variables

aleatorias. En particular, en el caso de dos variables aleatorias discretas X y Y , con conjuntos de valores respectivos $\{a_i\}$ y $\{b_j\}$, la independencia de estas dos variables se expresa en la siguiente forma

$$P(X^{-1}(\{a_i\}) \cap Y^{-1}(\{b_j\})) = P((X = a_i) \cap (Y = b_j)) = P(X = a_i)P(Y = b_j),$$

para todos $i, j \geq 1$.

Señalemos, finalmente en esta sección, que un suceso A y una variable aleatoria X se dice que son independientes, si A y B son independientes cualquiera sea $B \in \sigma(X)$.

10.6. Operaciones con variables aleatorias

Consideremos dos espacios medibles (Ω, \mathcal{F}) y (Ω', \mathcal{F}') , y una función $g : \Omega \rightarrow \Omega'$ tal que

$$g^{-1}(A') \in \mathcal{F} \quad \text{para todo } A' \in \mathcal{F}'.$$

La función g se denomina $(\mathcal{F}, \mathcal{F}')$ -**medible**. De acuerdo con esto, una variable aleatoria X es una función $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ -medible, pero en este caso se dice, simplemente, que X es \mathcal{F} -medible.

Sea $\Omega = \Omega' = \mathbb{R}$. Una función $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ -medible se llama **función Boreliana**. De manera general se llama función Boreliana a toda función g tal que $g : B \rightarrow \mathbb{R}$ y g es $(\mathcal{B}(B), \mathcal{B}(\mathbb{R}))$ -medible, donde B es un Boreliano.

Teorema 2.

Si X es una variable aleatoria y g una función Boreliana, entonces la función compuesta $g(X)$ es una variable aleatoria.

Demostración:

Sea B un Boreliano. Entonces

$$(g(X))^{-1}(B) = X^{-1}(g^{-1}(B)).$$

Pero $g^{-1}(B) = B'$ es un Boreliano, por ser g una función Boreliana, y como X es una variable aleatoria, se tiene

$$X^{-1}(g^{-1}(B)) = X^{-1}(B') \in \mathcal{F}.$$

■

Ejemplo 10.7 La función parte entera $\lfloor x \rfloor$, donde x es un número real, es una función Boreliana. Por consiguiente, si X es una variable aleatoria, entonces $\lfloor X \rfloor$ también lo es. ■

Ejercicio 5.

Sean X y Y variables aleatorias definidas en el mismo espacio de probabilidad. Pruebe que si X y Y son independientes, y g_1 y g_2 son funciones Borelianas, entonces $g_1(X)$ y $g_2(Y)$ son variables aleatorias independientes.

El siguiente teorema simplifica, en muchos casos, la prueba de que una función dada g es $(\mathcal{F}, \mathcal{F}')$ -medible. En lugar de probar que la preimagen mediante g de cualquier elemento de \mathcal{F}' pertenece a la σ -álgebra \mathcal{F} , es suficiente probar que la preimagen mediante g de cualquier elemento de un sistema generador de \mathcal{F}' pertenece a \mathcal{F} .

Teorema 3.

Sean los espacios medibles (Ω, \mathcal{F}) y (Ω', \mathcal{F}') , y \mathcal{C}' un sistema generador de \mathcal{F}' . Entonces, para que una función $g : \Omega \rightarrow \Omega'$ sea $(\mathcal{F}, \mathcal{F}')$ -medible es necesario y suficiente que $g^{-1}(A') \in \mathcal{F}$ para todo $A' \in \mathcal{C}'$.

Demostración:

Necesidad. Supongamos que g es $(\mathcal{F}, \mathcal{F}')$ -medible. Puesto que $\mathcal{C}' \subset \mathcal{F}'$, entonces se tiene que $g^{-1}(A') \in \mathcal{F}$ para todo $A' \in \mathcal{C}'$.

Suficiencia. Supongamos que $g^{-1}(A') \in \mathcal{F}$ para $A' \in \mathcal{C}'$. Notemos que la colección

$$\mathcal{D} = \{A' \subset \Omega' : g^{-1}(A') \in \mathcal{F}\},$$

es no vacía. Por otra parte, como la operación preimagen de conjuntos conserva las operaciones con conjuntos, entonces \mathcal{D} constituye una σ -álgebra en Ω' . Pero $\mathcal{C}' \subset \mathcal{D}$, y como $\sigma(\mathcal{C}')$ y \mathcal{D} son σ -álgebras que contienen a \mathcal{C}' , siendo $\sigma(\mathcal{C}')$ la mínima σ -álgebra que contiene a \mathcal{C}' , entonces $\mathcal{F}' = \sigma(\mathcal{C}') \subset \mathcal{D}$. Por consiguiente, g es $(\mathcal{F}, \mathcal{F}')$ -medible. ■

Ejercicio 6.

Sean $g : \Omega \rightarrow \Omega'$ y \mathcal{F} una σ -álgebra en Ω . Pruebe que

$$\mathcal{D} = \{A' \subset \Omega' : g^{-1}(A') \in \mathcal{F}\}$$

es una σ -álgebra en Ω' .

Puesto que la colección de subconjuntos abiertos de \mathbb{R} es un sistema generador de la σ -álgebra de los Borelianos, y la preimagen mediante una función continua g de un subconjunto abierto de \mathbb{R} es también un subconjunto abierto de \mathbb{R} , entonces del Teorema 3 inmediatamente se deduce el siguiente resultado.

Proposición 4.

Una función continua $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función Boreliana.

Como una función continua es Boreliana, y la composición de una función Boreliana con una variable aleatoria X es también una variable aleatoria, entonces, por ejemplo, son variables aleatorias $|X|$, y X^α con $X > 0$ y $\alpha > 0$.

Observe que según el Teorema 3, en la verificación de que una función es variable aleatoria pueden usarse diferentes familias de intervalos generadoras de la σ -álgebra de Borel, tales como $\{(t, +\infty)\}$, $\{[t, +\infty)\}$, $\{(-\infty, t]\}$ y $\{(-\infty, t)\}$, donde t es un número real.

Ejemplo 10.8 Probemos que si X y Y son variables aleatorias definidas sobre un espacio medible (Ω, \mathcal{F}) , entonces $\max(X, Y)$ y $\min(X, Y)$ son variables aleatorias.

En efecto, la colección, $\{(-\infty, t]\}$, $t \in \mathbb{R}$, es un sistema generador de los Borelianos, luego para todo número real t se tiene que

$$(\max(X, Y) \leq t) = (X \leq t) \cap (Y \leq t) = X^{-1}((-\infty, t]) \cap Y^{-1}((-\infty, t]) \in \mathcal{F}.$$

Análogamente, la colección $\{[t, +\infty)\}$ es un sistema generador de los Borelianos, y entonces se obtiene

$$(\min(X, Y) \geq t) = (X \geq t) \cap (Y \geq t) = X^{-1}([t, +\infty)) \cap Y^{-1}([t, +\infty)) \in \mathcal{F}.$$

Proposición 5.

Sean X y Y variables aleatorias definidas sobre un espacio medible (Ω, \mathcal{F}) . Entonces $X \pm Y$ es una variable aleatoria.

Demostración:

Probemos que para cada número real t se cumple que $(X + Y < t) \in \mathcal{F}$.

Sea $\omega \in (X + Y < t)$. La desigualdad $X(\omega) + Y(\omega) < t$ se escribe en la forma equivalente, $X(\omega) < t - Y(\omega)$. Note ahora que existe un número racional r tal que $X(\omega) < r < t - Y(\omega)$, y que estas últimas desigualdades se cumplen si, y solo si $X(\omega) < r$ y $Y(\omega) < t - r$. Luego,

$$\omega \in \bigcup_{r \in \mathbb{Q}} (X < r) \cap (Y < t - r)$$

de donde se obtiene la inclusión,

$$(X + Y < t) \subset \bigcup_{r \in \mathbb{Q}} (X < r) \cap (Y < t - r).$$

Es fácil ver que también se cumple la inclusión en el sentido inverso y, por consiguiente,

$$(X + Y < t) = \bigcup_{r \in \mathbb{Q}} (X < r) \cap (Y < t - r).$$

Finalmente, como $(X < r) \in \mathcal{F}$ y $(Y < t - r) \in \mathcal{F}$, y la unión en la última igualdad es numerable, entonces $(X + Y < t) \in \mathcal{F}$. ■

Ejemplo 10.9 Sean X y Y variables aleatorias definidas sobre (Ω, \mathcal{F}) . Probemos que $(X \leq Y) \in \mathcal{F}$ y $(X = Y) \in \mathcal{F}$.

Por la Proposición 5, la diferencia $X - Y$ es una variable aleatoria, y como $(-\infty, 0]$ es un Boreliano, se tiene

$$(X \leq Y) = (X - Y \leq 0) \in \mathcal{F}.$$

También, $(X \geq Y) \in \mathcal{F}$, de donde

$$(X = Y) = (X \leq Y) \cap (X \geq Y) \in \mathcal{F}.$$

Ejercicio 7.

Sean X y Y variables aleatorias definidas sobre un espacio medible (Ω, \mathcal{F}) y c un número real. Pruebe que entonces también son variables aleatorias: cX , X^2 , $|X|$, XY y X/Y ($Y(\omega) \neq 0, \omega \in \Omega$).

Ejemplo 10.10 Probemos que si X es una variable aleatoria y $\{X_n\}$ una sucesión de variables aleatorias, ambas definidas sobre un espacio medible (Ω, \mathcal{F}) , entonces el conjunto de puntos de convergencia de $\{X_n\}$ a X es un elemento de \mathcal{F} . Note que

$$\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n>N} \left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| < \frac{1}{k} \right\},$$

y puesto que $|X_n - X|$ es una variable aleatoria para cada n , se tiene entonces que

$$\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\} \in \mathcal{F}.$$

■

Definición 9.

Dado un espacio de probabilidad (Ω, \mathcal{F}, P) , se dice que una propiedad se cumple **casi seguramente** en Ω , si se cumple para todos los puntos de Ω , excepto quizás en un subconjunto $A \in \mathcal{F}$ tal que $P(A) = 0$.

Definición 10.

Se dice que la sucesión de variables aleatorias $\{X_n\}$ **converge casi seguramente**, o *casi seguro*, a la variable aleatoria X , si

$$P\left(\omega \in \Omega : \lim_n X_n(\omega) \neq X(\omega)\right) = 0$$

o, de forma equivalente, si

$$P\left(\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\right) = 1.$$

Este tipo de convergencia se denota por $X_n \xrightarrow{\text{c.s.}} X$ o $X_n \rightarrow X$ c.s.

Un ejemplo muy simple de convergencia casi segura es el siguiente. Consideremos el espacio de probabilidad $([0, 1], \mathbb{B}([0, 1]), \lambda)$, y $X_n(\omega) = \omega^n$ para todo $\omega \in [0, 1]$. La sucesión de variables aleatorias $\{X_n\}$ converge casi seguramente a 0, ya que solo no converge a 0 en $\omega = 1$, pero $\lambda(\{1\}) = 0$.

Ejemplo 10.11 Sea el espacio de probabilidad (Ω, \mathcal{F}, P) . Probemos que si $X_n \rightarrow X$ c.s y g es una función continua, entonces $g(X_n) \rightarrow g(X)$ c.s. Por la continuidad de $g(x)$ se tiene

$$\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\} = C_1 \subset C_2 = \{\omega \in \Omega : g(X_n(\omega)) \rightarrow g(X(\omega))\},$$

luego $P(C_2) = 1$, puesto que $P(C_1) = 1$. ■

Variable aleatoria ampliada

Algunas operaciones sobre una sucesión de variables aleatorias $\{X_n\}$ pueden generar los valores $+\infty$ y $-\infty$; por ejemplo, $\lim_n X_n$ y $\sup_n X_n$. También en problemas prácticos surgen variables aleatorias con estos valores. Un ejemplo de ello es el paseo aleatorio simple (vea el Ejercicio 10 de la Sección 9.6). Recordemos que este se define como una sucesión $\{S_n\}$ tal que $S_0 = 0$ y $S_n = X_1 + \cdots + X_n$, $n \geq 1$, donde $\{X_i\}$ es una sucesión de variables independientes e igualmente distribuidas tal que $P(X_1 = 1) = p = 1 - P(X_1 = -1)$, $0 < p < 1$. La variable aleatoria S_n se puede interpretar como la posición en el instante n de una partícula que se mueve aleatoriamente en el conjunto de los números enteros, y la condición $S_0 = 0$ significa que en el instante inicial la posición de la partícula es 0. Intuitivamente está claro que si $p > 1/2$, la

variable aleatoria T_0 , que representa el tiempo que demora la partícula en regresar a la posición 0, podría tomar el valor $+\infty$. Puede probarse que en este caso, $P(T_0 = +\infty) > 0$.

Los comentarios del párrafo anterior evidencian la necesidad de ampliar la Definición 4 de variable aleatoria:

Sea el espacio medible (Ω, \mathcal{F}) . Una **variable aleatoria ampliada** es una función $X : \Omega \rightarrow \bar{\mathbb{R}}$ tal que

$$X^{-1}(B) \in \mathcal{F} \text{ para todo } B \in \mathcal{B}(\mathbb{R}), X^{-1}(\{+\infty\}) \in \mathcal{F} \text{ y } X^{-1}(\{-\infty\}) \in \mathcal{F}.$$

10.7. Problemas y ejercicios

1. Considere el modelo de Bernoulli con $n = 2$. Sea X_i la variable aleatoria indicadora de la ocurrencia de éxito en el experimento i , $i = 1, 2$, y sea $X = X_1 + X_2$. Halle $\sigma(X_i)$, $i = 1, 2$, $\sigma(X)$ y $\sigma(\sigma(X_1) \cup \sigma(X_2))$.
2. Sean las variables aleatorias $\theta \sim U(0, 2\pi)$, $X = \cos \theta$ y $Y = \sin \theta$. Analice la independencia de los sucesos $A = (X < 0)$ y $B = (Y > 0)$.
3. Ilustre con un ejemplo que dos variables aleatorias sobre (Ω, \mathcal{F}) pueden ser independientes de acuerdo a una probabilidad P_1 y dependientes de acuerdo a otra probabilidad P_2 .
4. Sea $\{X_n\}$ una sucesión de variables aleatorias definidas sobre un espacio medible (Ω, \mathcal{F}) . Pruebe que

$$\sup_n X_n, \quad \inf_n X_n, \quad \limsup_n X_n, \quad \text{y} \quad \liminf_n X_n$$

son variables aleatorias, y que también lo es $\lim_n X_n$ si este límite existe, finito o infinito, para cada $\omega \in \Omega$.

Pruebe además, que

$$\{\omega \in \Omega : \limsup_n X_n(\omega) = \liminf_n X_n(\omega)\} \in \mathcal{F}.$$

5. Considere el espacio de probabilidad (Ω, \mathcal{F}, P) . Sea la sucesión de variables aleatorias $\{X_n\}$.
 - a) Analice la convergencia casi seguramente de $\{X_n\}$ si

$$P(X_n = 0) = 1/n^2 = 1 - P(X_n = 1), \quad n \geq 1.$$

b) Sean X_1, X_2, \dots , independientes y

$$P(X_n = 0) = 1/n = 1 - P(X_n = 1), \quad n \geq 1.$$

Pruebe que la probabilidad de que $\{X_n\}$ converja es igual a 0.

Sugerencia: Use el lema de Borel-Cantelli.

6. Considere el espacio de probabilidad (Ω, \mathcal{F}, P) . Sean X, Y y Z variables aleatorias tales que $X = Y$ c.s y $Y = Z$ c.s. Pruebe que entonces $X = Z$ c.s.
7. Considere el espacio de probabilidad (Ω, \mathcal{F}, P) . Sea la sucesión de variables aleatorias $\{X_n\}$ tal que $X_n \rightarrow X$ c.s y $X_n \rightarrow Y$ c.s. Pruebe que entonces $X = Y$ c.s.

VALOR ESPERADO

En este capítulo se define el valor esperado de una variable aleatoria arbitraria. Se considerará un espacio muestral Ω arbitrario, y las variables aleatorias definidas sobre el espacio de probabilidad (Ω, \mathcal{F}, P) , excepto cuando se indique algún espacio de probabilidad específico.

11.1. Variable aleatoria simple

Definición 1.

Se dice que una **variable aleatoria** es **simple**, si el conjunto de sus valores es finito.

Claro que toda variable aleatoria simple es una variable aleatoria discreta.

Sea X una variable aleatoria simple, y $\{a_1, a_2, \dots, a_n\}$ el conjunto de sus valores. Entonces

$$X^{-1}(\{a_i\}) = A_i \in \mathcal{F}, \quad i = 1, 2, \dots, n,$$

y la colección $\{A_i\}_{i=1}^n$ constituye una partición de Ω . Note que

$$X = \sum_{i=1}^n a_i I_{A_i}. \quad (11.1)$$

La representación (11.1) de una variable aleatoria X se denomina **representación estándar**.

Consideremos ahora una partición finita arbitraria de Ω , $\{B_j\}_{j=1}^m \subset \mathcal{F}$, y sea $X : \Omega \rightarrow \mathbb{R}$ una función definida como $X(\omega) = b_j$ si $\omega \in B_j$, $j = 1, 2, \dots, m$, donde los valores b_j no necesariamente son todos diferentes entre sí. Entonces X es una variable aleatoria simple que puede representarse en la forma,

$$X = \sum_{j=1}^m b_j I_{B_j}.$$

Luego una variable aleatoria simple puede tener diversas representaciones a través de diferentes particiones finitas de Ω en conjuntos medibles.

Definición 2.

Dada una partición finita de Ω , $\{B_j\}_{j=1}^m \subset \mathcal{F}$, y una variable aleatoria simple $X = \sum_{j=1}^m b_j I_{B_j}$, se denomina **valor esperado** (o **esperanza matemática**) de X , y se denota por EX , a la suma

$$EX = \sum_{j=1}^m b_j P(B_j).$$

Con el siguiente ejercicio se prueba que la definición de valor esperado para una variable aleatoria simple es correcta, pues este no depende de la partición del espacio muestral utilizada para representar la variable aleatoria.

Ejercicio 1.

Pruebe que si $\sum_{i=1}^l c_i I_{C_i}$ y $\sum_{j=1}^m b_j I_{B_j}$ son representaciones de una misma variable aleatoria simple, entonces

$$\sum_{i=1}^l c_i P(C_i) = \sum_{j=1}^m b_j P(B_j).$$

El valor esperado de una variable aleatoria se define primero, como ya se hizo, para una variable aleatoria simple, después para una variable aleatoria no negativa con base a la definición anterior, y por último, para una variable aleatoria arbitraria a partir de la definición para variable no negativa.

La definición de valor esperado de una variable aleatoria X es en realidad la definición de una integral; esta es la integral de la función medible (\mathcal{F} -medible) X , con respecto a la medida de probabilidad P . La notación en la forma de integral, del valor esperado EX de una variable aleatoria X , es la siguiente:

$$EX = \int_{\Omega} X dP \quad \text{o} \quad EX = \int_{\Omega} X(\omega) dP(\omega).$$

Entonces si, por ejemplo, X es una variable aleatoria simple, con representación estándar (11.1), se tiene

$$EX = \int_{\Omega} X dP = \sum_{i=1}^n a_i P(A_i).$$

Ejemplo 11.1 Considere A un conjunto medible. Como la función indicadora I_A se representa en la forma $1 \cdot I_A + 0 \cdot I_{A^c}$, entonces se obtiene,

$$EI_A = \int_{\Omega} I_A dP = P(A).$$

■

Es obvio que también son variables aleatorias simples la suma y el producto de un número finito de variables aleatorias simples, el cociente de dos variables aleatorias simples cuando este está definido, y toda función real compuesta con una variable aleatoria simple.

El siguiente teorema establece la propiedad de linealidad del valor esperado para variables aleatorias simples.

Teorema 1.

Sean X y Y variables aleatorias simples y a un número real. Entonces

1. $E(aX) = aEX$
2. $E(X + Y) = EX + EY$.

Demostración:

La demostración de la primera propiedad es obvia. Probemos la segunda. Expresemos las variables aleatorias X y Y por medio de las representaciones siguientes,

$$X = \sum_{i=1}^n a_i I_{A_i} \quad \text{y} \quad Y = \sum_{j=1}^m b_j I_{B_j}.$$

Entonces la variable aleatoria $X + Y$ se representa,

$$X + Y = \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) I_{A_i \cap B_j},$$

y su valor esperado se calcula,

$$\begin{aligned} E(X + Y) &= \sum_{i=1}^n \sum_{j=1}^m (a_i + b_j) P(A_i \cap B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m a_i P(A_i \cap B_j) + \sum_{i=1}^n \sum_{j=1}^m b_j P(A_i \cap B_j) \\ &= \sum_{i=1}^n a_i P(A_i) + \sum_{j=1}^m b_j P(B_j) = EX + EY. \end{aligned}$$

■

Ejemplo 11.2 Consideremos el espacio de probabilidad $([0, 1], \mathcal{B}([0, 1]), \lambda)$, y sea la función $X = aI_A + bI_B + cI_C$, donde $A = \mathbb{Q} \cap [0, 1]$, $B = \{1/\sqrt{3}\}$,

$C = [0, 1] \setminus (A \cup B)$, y a , b y c son números reales. La función X es una variable aleatoria simple, y

$$EX = \int_{[0,1]} X d\lambda = c.$$

Observe que la integral de la función X sobre el intervalo $[0, 1]$ no está definida en el sentido de la integral de Riemann. ■

Ejercicio 2.

Pruebe que si X y Y son variables aleatorias simples e independientes, entonces

$$EXY = EXEY.$$

11.2. Variable aleatoria no negativa

Definición 3.

Sea X una variable aleatoria no negativa. Se denomina **valor esperado** de X a

$$EX = \sup\{EZ : Z - \text{variable aleatoria simple y } Z \leq X\}.$$

Evidentemente, el valor esperado de una variable aleatoria simple, no negativa, satisface la definición general de valor esperado de una variable aleatoria no negativa.

El valor esperado de una variable aleatoria no negativa siempre existe, finito o infinito; en el caso en que este es finito, se dice que la variable aleatoria es **integrable**.

Note que la definición de valor esperado de la Sección 9.2 para una variable aleatoria no negativa X , definida en un espacio muestral discreto Ω , cumple con la definición general de valor esperado de esta sección. En efecto, si los valores de X son a_1, a_2, \dots , denotando, $A_i = X^{-1}(\{a_i\})$, se tiene que

$X = \sum_{i=1}^{\infty} a_i I_{A_i}$ y, según la definición de la Sección 9.2,

$$EX = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{i=1}^{\infty} a_i P(A_i).$$

Si ahora definimos la variable aleatoria simple, $Z_n = \sum_{i=1}^n a_i I_{A_i}$, entonces $Z_n \leq X$, $n \geq 1$, y $\sup_n EZ_n = EX$.

De la definición de valor esperado para variables no negativas inmediatamente se deduce la siguiente proposición.

Proposición 2.

Sean X y Y variables aleatorias no negativas. Si $X \leq Y$, entonces $EX \leq EY$.

Es de gran utilidad obtener condiciones bajo las cuales se cumple que

$$\lim_n EX_n = E(\lim_n X_n),$$

lo cual en la notación de integral se escribe

$$\lim_n \int_{\Omega} X_n dP = \int_{\Omega} \lim_n X_n dP.$$

El siguiente ejemplo muestra que este intercambio entre límite e integral no siempre es posible.

Ejemplo 11.3 Sean el espacio de probabilidad $((0, 1), \mathcal{B}((0, 1)), \lambda)$, y la sucesión de variables aleatorias $X_n = nI_{(0, 1/n)}$, $n \geq 1$, definidas en este espacio. Esta sucesión converge puntualmente a 0, pero $EX_n = 1$, luego,

$$\lim_n \int_{\Omega} X_n dP = \lim_n EX_n = 1 \neq 0 = E(\lim_n X_n) = \int_{\Omega} \lim_n X_n dP.$$

■

El siguiente teorema, denominado teorema de la convergencia monótona (TCM), establece una condición suficiente para el intercambio entre límite y valor esperado y tiene una importancia vital.

Teorema 3 (Teorema de la Convergencia Monótona. TCM).

Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $0 \leq X_n \uparrow X$.
Entonces

$$\lim_n EX_n = EX.$$

Demostración:

De las desigualdades, $0 \leq X_1 \leq X_2 \leq \dots \leq X$, se tiene que $EX_1 \leq EX_2 \leq \dots \leq EX$. Luego, $\lim_n EX_n$ existe y $\lim_n EX_n \leq EX$.

Probemos ahora que $\lim_n EX_n \geq EX$. Sea Z una variable aleatoria simple arbitraria tal que $Z \leq X$. Consideremos la representación estándar de esta variable simple, $Z = \sum_{i=1}^m c_i I_{A_i}$. Entonces $c_i \leq X(\omega)$ para todo $\omega \in A_i$. Dado $\epsilon > 0$, denotemos

$$A_{i,n} = \{\omega \in A_i : X_n(\omega) \geq c_i - \epsilon\}.$$

Puesto que la sucesión $\{X_n\}$ es creciente, se cumple que $A_{i,n} \subset A_{i,n+1}$, y por consiguiente, $\lim_n A_{i,n} = \cup_n A_{i,n}$. Verifiquemos que $\cup_n A_{i,n} = A_i$.

Como $A_{i,n} \subset A_i$ para todo $n \geq 1$, se tiene que $\cup_n A_{i,n} \subset A_i$. Si se supone ahora que $\omega \in A_i$, entonces como $X(\omega) \geq c_i$ y $X_n \uparrow X$, existe un valor $n \geq 1$ tal que $X_n(\omega) \geq c_i - \epsilon$, luego $\omega \in A_{i,n}$, de donde $\omega \in \cup_n A_{i,n}$. Así se obtiene que $\cup_n A_{i,n} = A_i$.

La sucesión $\{A_i\}_{i=1}^m$ es una partición de Ω , luego $A_{i,n} \cap A_{j,n} = \emptyset$ para $i \neq j$. Denotemos $B = (\cup_{i=1}^m A_{i,n})^c$. Se cumple que

$$X_n \geq \sum_{i=1}^m (c_i - \epsilon) I_{A_{i,n}} + 0 I_B,$$

de donde

$$EX_n \geq \sum_{i=1}^m (c_i - \epsilon) P(A_{i,n}),$$

y por la continuidad de la probabilidad se halla que

$$\lim_n EX_n \geq \sum_{i=1}^m (c_i - \epsilon) P(A_i) = EZ - \epsilon.$$

Entonces, puesto que ϵ es arbitrario, se tiene que $\lim_n EX_n \geq EZ$, y por la definición de EX se halla que $\lim_n EX_n \geq EX$. Finalmente, $\lim_n EX_n = EX$. ■

La siguiente proposición, cuya prueba se propone en el Ejercicio 2 de la Sección 12.7, muestra que para toda variable aleatoria no negativa X , siempre existe una sucesión de variables aleatorias simples que converge a X de manera monótona no decreciente.

Proposición 4.

Sea X una variable aleatoria no negativa arbitraria. La sucesión $\{X_n\}$, donde

$$X_n = \min \left(n, \frac{\lfloor 2^n X \rfloor}{2^n} \right), \quad n \geq 1,$$

y $\lfloor a \rfloor$ denota la parte entera de a , es una sucesión de variables aleatorias simples tal que $0 \leq X_n \uparrow X$. Si X está acotada, entonces la convergencia es uniforme.

La proposición anterior es válida también, si se sustituye, en el contexto correspondiente, la variable aleatoria X por una función Boreliana no negativa g . En este caso en lugar de la sucesión $\{X_n\}$ existe una sucesión de funciones Borelianas simples $\{g_n\}$ tal que $0 \leq g_n \uparrow g$.

A continuación, bajo ciertas condiciones, se establece la propiedad de linealidad del valor esperado para variables aleatorias no negativas.

Teorema 5.

Sean X y Y variables aleatorias no negativas y a un número real no negativo. Entonces

1. $E(aX) = aEX$
2. $E(X + Y) = EX + EY$.

Demostración:

Por la Proposición 4, existen sucesiones $\{X_n\}$ y $\{Y_n\}$ de variables aleatorias simples tales que $0 \leq X_n \uparrow X$ y $0 \leq Y_n \uparrow Y$. Luego, $0 \leq aX_n \uparrow aX$ y $0 \leq X_n + Y_n \uparrow X + Y$. Por lo tanto:

1. $E(aX) = \lim_n E(aX_n) = a \lim_n EX_n = aEX$,
2. $E(X + Y) = \lim_n E(X_n + Y_n) = \lim_n EX_n + \lim_n EY_n = EX + EY$,

donde tanto en la parte 1, como en la 2, la primera y última igualdades se cumplen por el TCM, y la segunda por la linealidad del valor esperado para variables aleatorias simples. ■

Teorema 6 (Corolario del TCM).

Sea $\{X_n\}$ una sucesión de variables no negativas. Entonces

$$E \sum_{i=1}^{\infty} X_i = \sum_{i=1}^{\infty} EX_i.$$

Demostración:

Denotemos $Y_n = \sum_{i=1}^n X_i$. Observe que $0 \leq Y_n \uparrow \sum_{i=1}^{\infty} X_i$. Luego,

$$E \sum_{i=1}^{\infty} X_i = E(\lim_n Y_n) = \lim_n EY_n = \lim_n \sum_{i=1}^n EX_i = \sum_{i=1}^{\infty} EX_i,$$

donde en la segunda igualdad se utilizó el TCM, y en la tercera la linealidad del valor esperado para variables aleatorias no negativas. ■

Ejemplo 11.4 Si X es una variable aleatoria con valores enteros no negativos, entonces, como es conocido (vea el Ejercicio 8 de la Sección 9.6),

$$EX = \sum_{k=1}^{\infty} P(X \geq k).$$

Este resultado se obtiene ahora aplicando el corolario del TCM a la siguiente identidad,

$$X = \sum_{k=1}^{\infty} I_{(X \geq k)}.$$

**Ejercicio 3.**

Sean X y Y variables aleatorias no negativas. Pruebe que si X y Y son independientes, entonces

$$EXY = EXEY.$$

La desigualdad que se establece en el siguiente teorema es de gran importancia en la teoría de probabilidades.

Teorema 7 (Desigualdad de Markov).

Sea X una variable aleatoria no negativa. Entonces para todo $\epsilon > 0$ se tiene que

$$P(X \geq \epsilon) \leq \frac{EX}{\epsilon}.$$

Demostración:

Como $X \geq XI_{(X \geq \epsilon)} \geq \epsilon I_{(X \geq \epsilon)}$, entonces $EX \geq \epsilon P(X \geq \epsilon)$, de donde se obtiene el resultado. ■

Ejemplo 11.5 Probemos que si $E|X| < \infty$, entonces $|X| < \infty$ c.s.

$$P(|X| = \infty) = P\left(\bigcap_{n=1}^{\infty} (|X| \geq n)\right) = \lim_n P(|X| \geq n) \leq \lim_n \frac{E|X|}{n} = 0.$$

En la segunda igualdad se aplicó la continuidad de la probabilidad, y al final, la desigualdad de Markov. ■

Ejercicio 4.

Sea $X \geq 0$. Pruebe que si $EX = 0$, entonces $X = 0$ c.s.

11.3. Variable aleatoria arbitraria

Sea X una variable aleatoria arbitraria. Denotemos, como en el caso del espacio muestral discreto,

$$X^+ = \max(X, 0) = XI_{(X \geq 0)} \quad \text{y} \quad X^- = -\min(X, 0) = -XI_{(X \leq 0)}.$$

Entonces, como ya es conocido, $X = X^+ - X^-$ y $|X| = X^+ + X^-$.

A continuación se define el valor esperado para una variable aleatoria arbitraria. La definición tiene la misma forma que la definición dada en el caso de un espacio muestral discreto.

Definición 4.

Sea X una variable aleatoria. Si $EX^+ < \infty$ ó $EX^- < \infty$, se define

$$EX = EX^+ - EX^-.$$

Cuando $EX^+ < \infty$ y $EX^- < \infty$ se dice que X es **integrable**. Si uno de los valores EX^+ o EX^- es finito y el otro infinito, se dice que el valor esperado es infinito. Cuando $EX^+ = EX^- = \infty$, el valor esperado no existe.

Por supuesto que una variable aleatoria X es integrable si, y solo si, es absolutamente integrable.

Teorema 8.

Sean X y Y variables aleatorias integrables, y a un número real. Entonces las variables aleatorias aX y $X+Y$ son integrables y se cumple:

1. $E(aX) = aEX$
2. $E(X + Y) = EX + EY$.

Demostración:

Que aX y $X + Y$ son integrables se deduce, respectivamente, de la igualdad

$|aX| = |a||X|$ y la desigualdad $|X + Y| \leq |X| + |Y|$, usando propiedades del valor esperado para variables aleatorias no negativas.

1. $aX = (aX)^+ - (aX)^-$. Si $a < 0$ se tiene

$$(aX)^+ = -aX^- \quad \text{y} \quad (aX)^- = -aX^+.$$

Entonces

$$E(aX) = E(-aX^-) - E(-aX^+) = -a(EX^- - EX^+) = aEX.$$

Si $a > 0$ se procede de forma análoga.

2. De las igualdades,

$$X + Y = (X + Y)^+ - (X + Y)^- \quad \text{y} \quad X + Y = X^+ - X^- + Y^+ - Y^-,$$

se deduce que

$$(X + Y)^+ + X^- + Y^- = (X + Y)^- + X^+ + Y^+.$$

Puesto que ambos miembros de la última igualdad y sus sumandos son variables aleatorias no negativas, se les halla sus valores esperados, y después de reagrupar los términos se obtiene

$$E(X + Y)^+ - E(X + Y)^- = EX^+ - EX^- + EY^+ - EY^-,$$

es decir, $E(X + Y) = EX + EY$. ■

Ejercicio 5.

Sean X y Y variables aleatorias integrables. Pruebe que:

1. $|EX| \leq E|X|$.
2. Si $X \leq Y$, entonces $EX \leq EY$.

En el siguiente ejercicio se establece una desigualdad importante en la teoría de probabilidades.

Ejercicio 6.

*(Desigualdad de Jensen). Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función convexa. Suponga que X es una variable aleatoria integrable y que $Eg(X)$ existe. Demuestre que entonces

$$g(EX) \leq Eg(X).$$

Ejercicio 7.

Sean X y Y variables aleatorias integrables. Pruebe que si X y Y son independientes, entonces

$$EXY = EXEY.$$

El siguiente ejemplo muestra que el recíproco de la propiedad establecida en el ejercicio anterior no se cumple.

Ejemplo 11.6 Dos variables aleatorias X y Y pueden ser dependientes y cumplir la igualdad, $EXY = EXEY$. Por ejemplo, sean X y X_1 independientes, $EX = EX_1 = 0$ y $Y = XX_1$. Es evidente que Y y X son dependientes y, por otro lado, $EXY = EXEY$, pues $EXY = EX^2X_1 = EX^2EX_1 = 0$ y $EXEY = 0$. ■

Definición 5.

Dada una variable aleatoria X y un conjunto medible A la integral sobre A de X con respecto a la medida P , denotada por $\int_A X dP$, se define,

$$\int_A X dP = \int_{\Omega} I_A X dP = E(I_A X).$$

De la definición anterior, en particular, se tiene que

$$\int_A dP = \int_{\Omega} I_A dP = EI_A = P(A).$$

Ejercicio 8.

Sea una variable aleatoria X no negativa definida sobre el espacio de probabilidad (Ω, \mathcal{F}, P) . Pruebe que la función μ definida como,

$$\mu(A) = \int_A X dP, \quad A \in \mathcal{F}$$

es una medida sobre \mathcal{F} .

La construcción hecha de la integral sobre Ω de una variable aleatoria con respecto a una medida de probabilidad, o sea, del valor esperado, es válida para cualquier medida μ cuando se integran funciones definidas sobre un conjunto de medida finita. Por ejemplo, si $g : B \rightarrow \mathbb{R}$ es una función Boreliana y $\lambda(B) < \infty$, la integral de g con respecto a la medida de Lebesgue λ se define de igual forma que lo hemos hecho para definir el valor esperado de una variable aleatoria. En este caso la variable aleatoria X se sustituye por g y las sucesiones de variables aleatorias $\{X_n\}$ por sucesiones de funciones Borelianas $\{g_n\}$.

Note que, a diferencia de la integral de Riemann, en una integral con respecto a una medida el espacio sobre el cual se integra puede ser de tal naturaleza que los conceptos de métrica y de continuidad pudieran no tener sentido.

11.4. Fórmula de cambio de variable

Teorema 9 (Fórmula de cambio de variable).

Sean X una variable aleatoria y g una función Boreliana. Entonces

$$Eg(X) = \int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) dP_X(x),$$

si una de las integrales existe.

Demostración:

a) Supongamos que g es una función Boreliana simple. Consideremos prime-

ro $g = I_C$, donde C es un conjunto Boreliano. Note que

$$g(X(\omega)) = I_C(X(\omega)) = I_{X^{-1}(C)}(\omega).$$

Luego,

$$\int_{\Omega} g(X) dP = \int_{\Omega} I_{X^{-1}(C)} dP = P(X^{-1}(C)) = P_X(C) = \int_{\mathbb{R}} I_C(x) dP_X.$$

Por la linealidad del valor esperado (integral) el teorema se cumple en general cuando g es simple.

b) Sea ahora $g \geq 0$ una función Boreliana. Entonces existe una sucesión de funciones Borelianas simples $\{g_n\}$ tal que $0 \leq g_n \uparrow g$. De donde, $0 \leq g_n(X) \uparrow g(X)$. Entonces,

$$\int_{\Omega} g(X) dP = \lim_n \int_{\Omega} g_n(X) dP = \lim_n \int_{\mathbb{R}} g_n(x) dP_X = \int_{\mathbb{R}} g(x) dP_X,$$

donde en la primera igualdad se utilizó el TCM, la segunda igualdad se cumple por la parte a) de la demostración, y la última por el TCM.

c) Sea g una función Boreliana arbitraria. Si $Eg(X)$ existe, entonces utilizando la parte b) de la demostración, se tiene que

$$Eg(X) = Eg^+(X) - Eg^-(X) = \int_{\mathbb{R}} g^+(x) dP_X - \int_{\mathbb{R}} g^-(x) dP_X = \int_{\mathbb{R}} g(x) dP_X.$$

■

Del teorema anterior se deduce que el valor esperado no depende de la forma de la función $X(\omega)$, sino solo de la distribución de probabilidad P_X .

Observación 7 El miembro izquierdo de la fórmula de cambio de variable del Teorema 9 es el valor esperado de $g(X)$ en el espacio de probabilidad (Ω, \mathcal{F}, P) y el miembro derecho es este valor esperado en el espacio de probabilidad $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$. Recordemos que a veces se hace referencia al valor esperado de una variable aleatoria X como el valor esperado de la distribución de esta variable aleatoria. ▲

DENSIDAD DE PROBABILIDAD

En este capítulo se introduce el concepto de densidad de probabilidad, que constituye una forma de representar la distribución de probabilidad de un tipo de variables aleatorias. Desde el punto de vista práctico los dos tipos fundamentales de variables aleatorias son, las discretas, y las que tienen densidad de probabilidad. Se tratan además algunas distribuciones de este último tipo que tienen gran interés teórico y práctico. En el capítulo se introducen también, de manera general para todo tipo de distribuciones, los conceptos de varianza y covarianza.

12.1. Integrales de Riemann y de Lebesgue

Cuando se integra con respecto a una medida, frecuentemente es necesario integrar funciones sobre conjuntos de medida infinita, tal como sucede a veces con la medida de Lebesgue.

Sea el espacio de medida $(\Omega, \mathcal{F}, \mu)$. Una medida μ se dice que es σ -finita, si existe una colección infinita numerable $\{A_i\}$ de conjuntos medibles tal que $\Omega = \bigcup_{i=1}^{\infty} A_i$ y $\mu(A_i) < \infty$ para todo $i \in \mathbb{N}$. La medida de Lebesgue λ , definida sobre el espacio medible $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ es σ -finita, pues, por ejemplo, si $A_i = (-i, i)$, $i \in \mathbb{N}$, entonces $\mathbb{R} = \bigcup_i A_i$, y $\lambda((-i, i)) = 2i < \infty$.

Si $\mu(\Omega) = \infty$, pero μ es σ -finita, entonces para una función f , \mathcal{F} -medible,

se define

$$\int_{\Omega} f d\mu = \lim_n \int_{A_n} f d\mu,$$

donde $A_n \uparrow \Omega$ y $\mu(A_i) < \infty, i \in \mathbb{N}$.

En particular, para la integral de una función Boreliana g con respecto a la medida de Lebesgue se tiene que

$$\int_{\mathbb{R}} g d\lambda = \lim_n \int_{(-n,n)} g d\lambda.$$

En el siguiente teorema, que no demostraremos, se establece la relación entre las integrales en el sentido de Riemann, propia e impropia, y las correspondientes integrales con respecto a la medida de Lebesgue para funciones de una variable real. Existen resultados análogos para integrales múltiples de funciones de varias variables.

Teorema 1.

Si una función $g : [a, b] \rightarrow \mathbb{R}$ es integrable según Riemann, entonces

$$\int_{[a,b]} g(x) d\lambda = \int_a^b g(x) dx.$$

Si una función $g : \mathbb{R} \rightarrow \mathbb{R}$ es absolutamente integrable en el sentido de Riemann, entonces

$$\int_{\mathbb{R}} g(x) d\lambda = \int_{-\infty}^{\infty} g(x) dx.$$

Puede demostrarse que una función es integrable según Riemann si, y solo si, el conjunto de puntos de discontinuidad de ésta tiene medida de Lebesgue 0. Sin embargo, una función puede ser integrable según Lebesgue siendo discontinua en todos los puntos.

En lo adelante, las funciones específicas que consideraremos en los cálculos serán integrables en el sentido de Riemann y, por lo tanto, sus integrales de Lebesgue y de Riemann coincidirán.

12.2. Densidad de probabilidad

Definición 1.

Se dice que la variable aleatoria X tiene **densidad de probabilidad**, si existe una función Boreliana no negativa $f(x)$ tal que

$$P_X(B) = \int_B f(x) d\lambda, \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}).$$

La función $f(x)$ se denomina densidad de probabilidad. También se dice que la distribución P_X tiene densidad de probabilidad.

Observe que

$$\int_{\mathbb{R}} f(x) d\lambda = P_X(\mathbb{R}) = 1.$$

Luego una densidad de probabilidad cumple dos propiedades análogas a las de la función de probabilidad de variables aleatorias discretas, las cuales establecemos en la siguiente proposición.

Proposición 2.

Toda densidad de probabilidad $f(x)$ cumple:

$$f(x) \geq 0, \quad x \in \mathbb{R}, \quad \text{y} \quad \int_{\mathbb{R}} f(x) d\lambda = 1.$$

Por otra parte, como los subconjuntos puntuales de \mathbb{R} tienen medida de Lebesgue 0, entonces para una variable aleatoria X con densidad de probabilidad se cumple que $P(X = a) = 0$ cualquiera sea el número real a . Por lo tanto, si $B = [a, b]$, cuando X tiene densidad de probabilidad $f(x)$ se tiene que

$$P_X(B) = P(a \leq X \leq b) = P(a < X < b) = \int_{(a,b)} f(x) d\lambda.$$

Ejemplo 12.1 Recordemos que al final de la Sección 10.4 se definió una variable aleatoria X con distribución uniforme en el intervalo (a, b) como aquella cuya distribución de probabilidad P_X se determina por la fórmula,

$$P_X(B) = \frac{\lambda(B \cap (a, b))}{b - a}, \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}).$$

Entonces

$$P_X(B) = \frac{1}{b - a} \int_{\mathbb{R}} I_{(B \cap (a, b))} d\lambda = \int_{\mathbb{R}} \frac{1}{b - a} I_B I_{(a, b)} d\lambda = \int_B \frac{1}{b - a} I_{(a, b)} d\lambda.$$

Luego la variable aleatoria X tiene densidad de probabilidad

$$f(x) = \frac{1}{b - a} I_{(a, b)}(x).$$

Además,

$$\int_{\mathbb{R}} f(x) d\lambda = 1.$$

Por otro lado, si $a < c < d < b$, entonces

$$\begin{aligned} P_X(B) &= P(c < X < d) = P(c \leq X \leq d) = \int_{[c, d]} f(x) d\lambda \\ &= \int_c^d \frac{1}{b - a} dx = \frac{d - c}{b - a}. \end{aligned}$$

Observe que las dos últimas integrales, según Lebesgue y según Riemann, coinciden. ■

De forma equivalente a la definición dada anteriormente, ahora se puede definir una variable aleatoria con distribución uniforme en el intervalo (a, b) como aquella cuya densidad de probabilidad $f(x)$ se expresa por la fórmula,

$$f(x) = \frac{1}{b - a} I_{(a, b)}(x).$$

En los cálculos numéricos aproximados, se puede asumir que el error por redondeo tiene distribución uniforme. Por ejemplo, si el calculo aproximado se expresa solo hasta la segunda cifra decimal, el error por redondeo tiene distribución uniforme en el intervalo $(-0,01, 0,01)$.

Observación 8 La densidad de probabilidad y la función de probabilidad son dos de las formas de expresar la distribución de probabilidad P_X de una variable aleatoria X . En muchas aplicaciones no es necesario conocer cómo está definida una variable aleatoria, basta conocer cuál es la distribución de probabilidad de la variable, expresada de acuerdo a la densidad de probabilidad o la función de probabilidad, según sea el caso. ▲

La técnica ya utilizada anteriormente en la demostración de varias propiedades, a saber, primero la demostración para el caso de variables aleatorias o funciones Borelianas simples; luego para no negativas; y finalmente para variables aleatorias o funciones Borelianas arbitrarias, es estándar, y se utilizará en la demostración del siguiente teorema.

Teorema 3.

Sean una variable aleatoria X con densidad de probabilidad f , y g una función Boreliana. Entonces

$$\int_{\mathbb{R}} g(x) dP_X = \int_{\mathbb{R}} g(x) f(x) d\lambda,$$

si una de las integrales existe.

Demostración:

a) Consideremos $g = I_B$, donde $B \in \mathcal{B}(\mathbb{R})$. Se tiene

$$\int_{\mathbb{R}} g(x) dP_X = \int_{\mathbb{R}} I_B(x) dP_X = P_X(B) = \int_B f(x) d\lambda = \int_{\mathbb{R}} I_B(x) f(x) d\lambda.$$

Luego, por la linealidad de la integral, el teorema es válido para funciones Borelianas simples.

b) Sea $g \geq 0$. Entonces existe una sucesión de funciones Borelianas simples $\{g_n\}$ tal que $0 \leq g_n \uparrow g$. De donde, $0 \leq g_n f \uparrow gf$. Por lo tanto,

$$\int_{\mathbb{R}} g(x) dP_X = \lim_n \int_{\mathbb{R}} g_n(x) dP_X = \lim_n \int_{\mathbb{R}} g_n(x) f(x) d\lambda = \int_{\mathbb{R}} g(x) f(x) d\lambda.$$

La primera y últimas igualdades de las igualdades anteriores se cumplen por el TCM, y la segunda es válida por la parte a) de la demostración.

c) En el caso de una función Boreliana arbitraria la demostración ya es evidente. ■

Utilizando los Teoremas 9 y 3 se concluye el siguiente resultado que formulamos en forma de una proposición

Proposición 4.

Sean X una variable aleatoria con densidad de probabilidad f , y g una función Boreliana. Entonces

$$\begin{aligned} Eg(X) &= \int_{\Omega} g(X(\omega))dP = \int_{\mathbb{R}} g(x)dP_X \\ &= \int_{\mathbb{R}} g(x)f(x)d\lambda = \int_{-\infty}^{\infty} g(x)f(x)dx, \end{aligned}$$

donde se ha supuesto que la última integral es absolutamente convergente, o que esta existe si la distribución está concentrada en un intervalo finito, digamos, $f(x) = 0$ si $x \notin [a, b]$.

Ejemplo 12.2 Sea X una variable aleatoria con distribución uniforme en el intervalo (a, b) . Entonces, si denotamos por $f(x)$ la densidad de probabilidad de X , se tiene

$$EX = \int_{\mathbb{R}} xf(x)d\lambda = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

Este resultado es intuitivamente obvio. ■

Ejemplo 12.3 Una variable aleatoria X se dice que tiene **distribución de Cauchy** estándar si tiene la siguiente densidad de probabilidad

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

La integral

$$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx,$$

no converge absolutamente. Entonces el valor esperado pudiera ser infinito o no existir. Puede verificarse que el valor esperado de X no existe. ■

Ejemplo 12.4 Se dice que una variable aleatoria X tiene **distribución exponencial** de parámetro λ , y se denota por $X \sim \text{Exp}(\lambda)$, si tiene la siguiente densidad de probabilidad

$$f(x) = \lambda e^{-\lambda x} I_{(0, \infty)}(x),$$

donde $\lambda > 0$. El valor esperado de X se calcula integrando por partes, de donde se halla que

$$EX = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

■

Señalemos finalmente en esta sección, que en las aplicaciones prácticas usualmente una densidad de probabilidad caracteriza la distribución de una variable aleatoria cuyo conjunto de valores es un intervalo, finito o infinito.

12.3. Varianza

Definición 2.

Se denomina **momento de orden n** de una variable aleatoria X al valor esperado EX^n , $n = 1, 2, \dots$, cuando este existe.

Ejemplo 12.5 a) Sea $X \sim U\{1, 2, \dots, N\}$. El momento de orden 2 de X se calcula,

$$EX^2 = \frac{1}{N}(1^2 + 2^2 + \dots + N^2) = \frac{(N+1)(2N+1)}{6}.$$

b) Si $X \sim \text{Bernoulli}(p)$, entonces $EX^n = p$, $n \geq 1$.

c) Sea $X \sim \text{Exp}(\lambda)$. Usando el método de integración por partes e inducción matemática es fácil calcular el momento de orden n de esta distribución:

$$EX^n = \int_0^{\infty} x^n \lambda e^{-\lambda x} dx = \frac{n!}{\lambda^n}, \quad n \geq 1.$$

■

Ejercicio 1.

Sean X una variable aleatoria y $1 \leq \alpha < \beta$. Pruebe que

$$\text{si } E|X|^\beta < \infty, \text{ entonces } E|X|^\alpha < \infty.$$

Del ejercicio anterior se obtiene, en particular, que si $EX^2 < \infty$, entonces $E|X| < \infty$.

Definición 3.

Sea una variable aleatoria X tal que $EX^2 < \infty$. La **varianza** de X , denotada por $V(X)$, se define

$$V(X) = E(X - EX)^2.$$

Se llama **desviación estándar** de la variable aleatoria X , y se denota por σ_X , a $\sigma_X = \sqrt{V(X)}$.

La varianza de una variable aleatoria X es una medida de la desviación de los valores de X con respecto a su valor esperado, a través del promedio ponderado de estas desviaciones al cuadrado de acuerdo a la distribución de probabilidad. Mientras más pequeña sea la varianza de una variable aleatoria, más concentrada estará su distribución de probabilidad alrededor del valor esperado.

Es evidente que si X es una variable aleatoria constante, es decir tal que $X(\omega) = c$, para todo $\omega \in \Omega$, donde c es un número real, entonces $V(X) = 0$.

Ejercicio 2.

Verifique que

$$V(X) \leq E(X - c)^2,$$

cualquiera sea el número real c .

El resultado del ejercicio anterior significa que la menor desviación al cuadrado promedio de los valores de X con respecto a una constante es su valor esperado.

En la siguiente proposición, la primera propiedad se obtiene inmediatamente de la definición de varianza, y la segunda aplicando la linealidad del valor esperado. La segunda propiedad brinda una forma alternativa, con respecto a la definición, para el cálculo de la varianza.

Proposición 5.

Sean X una variable aleatoria tal que $EX^2 < \infty$ y c un número real. Entonces se cumplen las siguientes propiedades:

1. $V(cX) = c^2V(X)$.
2. $V(X) = EX^2 - (EX)^2$.

Sean X una variable aleatoria, y $Z = \frac{X-EX}{\sigma_X}$. Es fácil verificar que entonces $EZ = 0$ y $V(Z) = 1$. El proceso que consiste en restar a una variable aleatoria X su valor esperado y dividir por su desviación estándar, se denomina estandarización de la variable. La variable aleatoria Z se denomina **variable estandarizada** a partir de la variable X .

Ejemplo 12.6 Vea el Ejemplo 12.5.

a) Si $X \sim \text{Bernoulli}(p)$, entonces

$$V(X) = EX^2 - (EX)^2 = p - p^2 = p(1 - p).$$

b) Sea $X \sim \text{Exp}(\lambda)$. Entonces

$$V(X) = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Para esta distribución se cumple que $\sigma_X = EX$. ■

Ejercicio 3.

Verifique que:

1. Si $X \sim U(a, b)$, entonces $V(X) = \frac{(b-a)^2}{12}$.
2. Si $X \sim \text{Poisson}(a)$, entonces $V(X) = a$.
3. Si $X \sim \text{Geo}(p)$, entonces $V(X) = \frac{1-p}{p^2}$.

12.4. Distribución normal

Se dice que una variable aleatoria X tiene **distribución normal** o de Gauss, con parámetros μ y σ , lo cual se denota $X \sim N(\mu, \sigma^2)$, si X tiene densidad de probabilidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

donde $\mu \in \mathbb{R}$ y $\sigma > 0$.

Sea $X \sim N(\mu, \sigma^2)$. No es difícil calcular el valor de la siguiente integral que define el valor esperado de X ,

$$EX = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \mu.$$

La varianza de X se expresa por la integral,

$$V(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx,$$

cuyo valor es $V(X) = \sigma^2$. Luego $\sigma_X = \sigma$.

Para la probabilidad $P(a < X < b)$ se halla que

$$\begin{aligned} P(a < X < b) &= \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \end{aligned} \tag{12.1}$$

donde

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

El integrando de la función $\Phi(x)$ es la densidad de probabilidad correspondiente a una variable aleatoria con distribución $N(0, 1)$. Una variable aleatoria con esta distribución se dice que tiene **distribución normal estándar**. Observe que $\Phi(0) = 1/2$, y $\Phi(-x) = 1 - \Phi(x)$ para $x > 0$.

Estas propiedades de la función $\Phi(x)$ posibilitan realizar el cálculo de la probabilidad de que una variable aleatoria con distribución normal tome valores en un intervalo dado, usando una tabla de valores aproximados de $\Phi(x)$ construida solo para valores de x mayores que 0. El cálculo de los valores de $\Phi(x)$ también está implementado en la mayoría de los software usuales.

Ejemplo 12.7 Sea $X \sim N(\mu, \sigma^2)$. Usando (12.1) se halla que

$$p = P(|X - \mu| < 3\sigma) = P(\mu - 3\sigma < X < \mu + 3\sigma) = 2\Phi(3) - 1 \approx 0,997.$$

En el lenguaje Python este cálculo se puede realizar con el siguiente fragmento de código, donde $\Phi(x) = \text{norm.cdf}(x)$, y *cdf* significa *cumulative distribution function*.

```
from scipy.stats import norm
p=2*norm.cdf(3)-1
```

El valor $p \approx 0,997$ se interpreta de la siguiente manera: si observamos valores de X , la gran mayoría de estos valores se encontrarán en el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$. ■

En muchas ocasiones se requiere calcular un valor de la función inversa de $\Phi(x)$, es decir, dada una probabilidad p se requiere calcular el valor de x tal que $\Phi(x) = p$. Con el siguiente código en lenguaje Python se realiza este cálculo de manera aproximada.

```
from scipy.stats import norm
x=norm.ppf(p)
```

Aquí *ppf* significa *probability point function*.

12.5. Distribuciones Gamma, Chi-cuadrado y Beta

Se dice que una variable aleatoria X tiene **distribución Gamma** con parámetros λ y α , lo cual se denota $X \sim \text{Gamma}(\lambda, \alpha)$, si X tiene densidad de probabilidad

$$f(x) = \frac{\lambda(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} I_{(0,\infty)}(x),$$

donde $\lambda > 0$, $\alpha > 0$, y $\Gamma(\alpha)$ es la función gamma, o sea,

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

En particular, si $\alpha = n$, donde n es un número natural, se tiene $\Gamma(n) = (n-1)!$.

Con la variación del parámetro λ no varía la forma del gráfico de la densidad de probabilidad $f(x)$, este es un parámetro de escala. La variación de α hace variar la forma del gráfico de $f(x)$. Por eso, a α y λ se les llama parámetro de forma y de escala, respectivamente.

Cuando $\alpha = 1$ se tiene la distribución exponencial $\text{Exp}(\lambda)$. Cuando $\lambda = 1/2$ y $\alpha = n/2$, es decir, si $X \sim \text{Gamma}(1/2, n/2)$, entonces se dice que X tiene distribución **chi-cuadrado con parámetro n** (o n grados de libertad), y se denota por $X \sim \chi^2(n)$.

Ejercicio 4.

Verifique que si $X \sim \text{Gamma}(\lambda, \alpha)$, entonces

$$EX = \frac{\alpha}{\lambda} \quad \text{y} \quad V(X) = \frac{\alpha}{\lambda^2}.$$

Del ejercicio anterior se halla que si $X \sim \chi^2(n)$, entonces $EX = n$ y $V(X) = 2n$.

Se dice que una variable aleatoria X tiene **distribución Beta** con parámetros a y b , lo cual se denota $X \sim \text{Beta}(a, b)$, si X tiene densidad de probabilidad

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{\mathcal{B}(a, b)} I_{(0,1)}(x),$$

donde $a > 0$ y $b > 0$, y $\mathcal{B}(a, b)$ es la función beta, es decir,

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

La función $\mathcal{B}(a, b)$ se puede representar como,

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

La distribución uniforme en el intervalo $(0, 1)$ es un caso particular de la distribución Beta cuando $a = b = 1$.

Ejercicio 5.

Verifique que si $X \sim \text{Beta}(a, b)$, entonces

$$EX = \frac{a}{a+b} \quad \text{y} \quad V(X) = \frac{ab}{(a+b+1)(a+b)^2}.$$

12.6. Covarianza y coeficiente de correlación

Antes de dar la próxima definición, notemos que de la desigualdad $2|XY| \leq X^2 + Y^2$ se deduce que si $EX^2 < \infty$ y $EY^2 < \infty$, entonces $E|XY| < \infty$.

Definición 4.

Sean $EX^2 < \infty$ y $EY^2 < \infty$. Se denomina **covarianza** de X y Y , y se denota por $\text{cov}(X, Y)$, a

$$\text{cov}(X, Y) = E(X - EX)(Y - EY).$$

Si $\text{cov}(X, Y) = 0$, entonces se dice que las variables aleatorias X y Y están **no correlacionadas**.

No es difícil ver que

$$\text{cov}(X, Y) = EXY - EXEY.$$

Por otro lado, si las variables aleatorias X y Y son independientes, entonces $EXY = EXEY$ y, por consiguiente, $\text{cov}(X, Y) = 0$. Luego la independencia de dos variables aleatorias implica su no correlación; sin embargo, como muestra el siguiente ejemplo, el recíproco de esta afirmación no se cumple (vea también el Ejemplo 11.6).

Ejemplo 12.8 Consideremos la variable aleatoria X con distribución

$$P(X = -1) = \frac{1}{4}, \quad P(X = 0) = \frac{1}{2}, \quad P(X = 1) = \frac{1}{4}.$$

Tomemos $Y = X^2$. La distribución de la variable aleatoria Y es

$$P(Y = 0) = \frac{1}{2} \quad P(Y = 1) = \frac{1}{2}.$$

Las variables aleatorias X y Y son dependientes, puesto que

$$P(X = 0, Y = 0) = \frac{1}{2} \neq \frac{1}{2} \cdot \frac{1}{2} = P(X = 0)P(Y = 0).$$

Además, $EX = 0$, $EY = \frac{1}{2}$ y $EXY = EX^3 = 0$. Por lo tanto, $\text{cov}(X, Y) = 0$, o sea, X y Y están no correlacionadas. ■

Proposición 6.

Sean las variables aleatorias X_1, X_2, \dots, X_n , tales que $EX_i^2 < \infty$, $i = 1, 2, \dots, n$. Entonces

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j).$$

Si las variables aleatorias X_1, X_2, \dots, X_n son independientes dos a dos, entonces

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i).$$

Demostración:

Probaremos el caso $n = 2$. La demostración general es una simple generalización de este caso. Se tiene que

$$\begin{aligned} V(X + Y) &= E(X + Y - E(X + Y))^2 \\ &= E[(X - EX)^2 + (Y - EY)^2 + 2(X - EX)(Y - EY)] \\ &= V(X) + V(Y) + 2\text{cov}(X, Y), \end{aligned}$$

lo cual demuestra la primera parte de la proposición. La segunda parte es evidente, puesto que si las variables aleatorias X_i y X_j , $i \neq j$, son independientes, entonces $\text{cov}(X_i, X_j) = 0$. ■

Ejemplo 12.9 Anteriormente se vio (en el Ejemplo 9.2), que una variable aleatoria X con distribución Binomial(n, p) puede representarse en la forma

$$X = X_1 + X_2 + \cdots + X_n,$$

donde $X_i \sim \text{Bernoulli}(p)$, $i = 1, 2, \dots, n$ y las variables aleatorias sumandos son independientes. Como $V(X_i) = p(1-p)$, $i = 1, 2, \dots, n$, entonces, por la Proposición 6 se tiene que

$$V(X) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = np(1-p).$$

■

Similar al caso de una variable aleatoria con distribución binomial, una variable aleatoria X con distribución Hipergeo(N, M, n) también puede representarse como la suma $X = \sum_{i=1}^n X_i$, donde las variables aleatorias X_i , $1 \leq i \leq n$, tienen distribución Bernoulli($\frac{M}{N}$). Sin embargo, en este caso no es posible hallar la varianza como se hizo en el ejemplo anterior, puesto que las variables aleatorias sumandos son dependientes; ahora se requiere el uso de la fórmula general para la varianza de una suma de variables aleatorias; es bastante laborioso realizar los cálculos y solo damos el resultado, $V(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$.

Definición 5.

Sean $EX^2 < \infty$ y $EY^2 < \infty$ y $\sigma_X \cdot \sigma_Y \neq 0$. Se denomina **coeficiente de correlación** de X y Y , y se denota por $\rho(X, Y)$, a

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

El coeficiente de correlación de dos variables aleatorias se denotará simplemente por la letra ρ cuando esté claro a qué variables aleatorias se refiere.

Proposición 7.

El coeficiente de correlación ρ de dos variables aleatorias X y Y cumple las siguientes desigualdades,

$$-1 \leq \rho \leq 1.$$

Demostración:

Sea α un número real. Usando propiedades de la varianza hallamos que

$$0 \leq V(\alpha X + Y) = \alpha^2 V(X) + 2\alpha \text{cov}(X, Y) + V(Y) = g(\alpha).$$

Pero $g(\alpha) \geq 0$ para todo $\alpha \in \mathbb{R}$, si, y solo si, el discriminante correspondiente al polinomio de segundo grado $g(\alpha)$ es menor o igual que 0, es decir,

$$(2\text{cov}(X, Y))^2 - 4V(X)V(Y) \leq 0,$$

desigualdad equivalente a la desigualdad $(\text{cov}(X, Y))^2 \leq V(X)V(Y)$, de donde se deduce que $-1 \leq \rho \leq 1$. ■

Ejercicio 6.

Sean X y Y variables aleatorias. Pruebe que si $Y = \alpha X + \beta$, donde α y β son números reales y $\alpha \neq 0$, entonces el coeficiente de correlación ρ de X y Y es igual a 1 o -1 .
(En realidad basta que $Y = \alpha X + \beta$ c.s para que ρ sea igual a 1 o -1).

En la Sección 14.2 se demuestra el recíproco de la propiedad enunciada en el ejercicio anterior: si $\rho = 1$ o $\rho = -1$, entonces $\alpha Y + \beta$ c.s.

El coeficiente de correlación es una medida de la relación lineal entre dos variables aleatorias. Cuando el valor de ρ está cerca de -1 o de 1 , esto significa que existe una relación lineal aproximada entre las dos variables aleatorias, y si el valor de ρ está lejos de estos valores, por ejemplo cerca de 0 , entonces la relación lineal entre las variables es muy débil o no existe. En próximos capítulos quedará más clara esta interpretación.

12.7. Problemas y ejercicios

1. Sean X y Y variables aleatorias simples. Pruebe las siguientes propiedades:

- a) Si $X = 0$ c.s, entonces $EX = 0$.
- b) Si $X \geq 0$ y $EX = 0$, entonces $X = 0$ c.s.
- c) Si $Y = X$ c.s, entonces $EY = EX$.

2. Pruebe la Proposición 4.

3. Sea X una variable aleatoria discreta con función de probabilidad $p_i = f(a_i)$, $i \geq 1$, y sea g una función Boreliana no negativa. Pruebe que entonces

$$E(g(X)) = \sum_{i=1}^{\infty} g(a_i)p_i.$$

4. Demuestre que si X una variable aleatoria tal que $X = 0$ c.s, entonces $EX = 0$.

5. Pruebe que:

- a) Si $P(A) = 0$, entonces $\int_A X dP = 0$.
- b) Cuando $X > 0$, si $\int_A X dP = 0$, entonces $P(A) = 0$.

6. Una variable aleatoria X tiene densidad de probabilidad

$$f(x) = \begin{cases} a + bx^2, & \text{si } 0 \leq x \leq 1 \\ 0, & \text{en otro caso,} \end{cases}$$

donde a y b son constantes reales. Halle a y b si $EX = 3/5$.

7. *Ausencia de Memoria.* Compruebe que una variable aleatoria X con distribución exponencial cumple:

$$P(X > s + t | X > t) = P(X > s), \quad \text{para todos } s, t \geq 0.$$

8. Sea X una variable aleatoria con densidad de probabilidad, y sea $g(x) = P(X > x)$ tal que $g(0) = 1$. Pruebe que la ecuación funcional

$$g(t + x) = g(t)g(x), \quad \text{para todos } x, y \geq 0,$$

tiene la única solución $g(x) = e^{-\lambda x}$, donde $\lambda > 0$. Explique la relación de este resultado con la propiedad de ausencia de memoria de la distribución exponencial.

9. Pruebe que si X es una variable aleatoria tal que $X = c$ c.s, donde c es una constante real, entonces $V(X) = 0$. Pruebe que, inversamente, si $V(X) = 0$, entonces $X = EX$ c.s.
10. * Sea X una variable aleatoria tal que $P(0 \leq X \leq c) = 1$, donde $c > 0$. Pruebe que

$$V(X) \leq \frac{c^2}{4}.$$

Sugerencia: Pruebe primero que $V(X) \leq c^2[\alpha(1 - \alpha)]$, donde $\alpha = \frac{EX}{c}$.

11. (Desigualdad de Lyapunov). Pruebe que si $0 < s < t$ y $E|X|^s < \infty$, entonces

$$(E|X|^s)^{1/s} \leq (E|X|^t)^{1/t}.$$

12. Use la desigualdad de Jensen para probar que la media aritmética de los números reales no negativos, x_1, x_2, \dots, x_n , es mayor o igual que su media geométrica.
13. a) Sea $X \sim N(12, 4)$. Calcule $P(X > 13)$ y $P(10 < X < 11)$. Halle el valor de c tal que $P(X > c) = 0,1$.
b) Sea Z una variable aleatoria con distribución normal estándar. Verifique que $P(|Z| > x) = 2P(Z > x)$, donde $x \geq 0$.
14. Sea $X \sim N(0, 1)$. Verifique que $EX^{n+1} = nEX^{n-1}$, $n \geq 1$. Halle todos los momentos de la variable aleatoria X .
15. Un mensaje binario x está codificado con los números -2 ó 2 . Cuando se realiza la transmisión, se recibe un mensaje $R = x + N$ que consiste en el mensaje de interés x mezclado en forma de suma con un ruido N que lo perturba. Suponga que N es una variable aleatoria con distribución normal estándar y que el mensaje recibido R se decodifica de acuerdo a la siguiente regla: Si $R \geq 0.5$ entonces se concluye que el mensaje x es 2 , en caso contrario se concluye que el mensaje x es -2 . Calcule la probabilidad de cada uno de los dos errores que pueden ocurrir en la decodificación.
16. Sea $F(x) = 1 - e^{-x^2}$ la función de distribución de la variable aleatoria X . Calcule: $P(X > 2)$, $P(1 < X < 3)$, EX y $V(X)$.
17. Verifique que si $X \sim \text{Binomial}(p, m)$, entonces $V(X) = \frac{m(1-p)}{p^2}$.

18. Sean $X_1 = \alpha_1 X + \beta_1$ y $Y_1 = \alpha_2 Y + \beta_2$, con $\alpha_1 \alpha_2 \neq 0$ y $\beta_1, \beta_2 \in \mathbb{R}$. Halle una expresión para $\rho(X_1, Y_1)$ en dependencia de $\rho(X, Y)$.

FUNCIÓN DE DISTRIBUCIÓN

En este capítulo se introduce el concepto de función de distribución de una variable aleatoria, el cual constituye una forma general de representar la distribución de cualquier tipo de variable aleatoria X , y es equivalente a la distribución de probabilidad P_X . Se establece además la relación entre la función de distribución y la densidad de probabilidad en el caso en que esta última exista.

13.1. Función de distribución

Definición 1.

Se denomina **función de distribución** de una variable aleatoria X a la función $F(x)$ definida como

$$F(x) = P_X((-\infty, x]) = P(X \leq x), \quad x \in \mathbb{R}.$$

Sea X una variable aleatoria discreta con función de probabilidad $p_i = f(a_i)$, $i = 1, 2, \dots, n$, y $a_1 < a_2 < \dots < a_n$. No es difícil ver que la función de

distribución de X se expresa:

$$F(x) = \begin{cases} 0, & \text{si } x < a_1 \\ \sum_{i=1}^j p_i & \text{si } a_j \leq x < a_{j+1}, \quad j = 1, \dots, n-1 \\ 1, & \text{si } x \geq a_n. \end{cases}$$

Si el conjunto de valores de la variable aleatoria discreta es infinito, su función de distribución se representa de manera similar. En el caso particular en que $P(X = c) = 1$, la función de distribución de X es $F(x) = I_{[c, +\infty)}(x)$.

Cuando la variable aleatoria X tiene densidad de probabilidad $f(x)$, entonces su función de distribución $F(x)$ se expresa (vea la Definición 1, Capítulo 12)

$$F(x) = P_X((-\infty, x]) = \int_{-\infty}^x f(t)dt. \quad (13.1)$$

Inversamente, si la función de distribución $F(x)$ de una variable aleatoria X se representa a través de una función $f(x)$ por la fórmula anterior, entonces $f(x)$ es la densidad de probabilidad de X . Por consiguiente, si existe la densidad de probabilidad de una variable aleatoria, sus funciones de distribución y de densidad se determinan mutuamente.

Observación 2.4. Naturalmente que la función de distribución de una variable aleatoria con densidad de probabilidad es una función continua. Sin embargo, una función de distribución puede ser continua y no ser representable por medio de una densidad de probabilidad; el análisis de esta situación queda fuera del alcance de este texto.

De (13.1) se obtiene la igualdad

$$\frac{dF(x)}{dx} = f(x),$$

en los puntos de continuidad de $f(x)$. Además,

$$P(a < X < b) = \int_a^b f(t)dt = F(b) - F(a).$$

Ejemplo 13.1 a) Sea $X \sim U(a, b)$. Entonces

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

En particular, si $U \sim U(0, 1)$, la función de distribución $F(x)$ de U queda definida como, $F(x) = x$ para $x \in (0, 1)$; $F(x) = 0$ si $x \leq 0$, y $F(x) = 1$ si $x \geq 1$.

b) Sea $X \sim \text{Exp}(\lambda)$. Entonces $F(x) = 0$ para $x \leq 0$, y si $x > 0$ se tiene

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

c) La función de distribución correspondiente a una variable aleatoria normal estándar se denota de manera especial, y es la función $\Phi(x)$ que ya se introdujo antes en la Sección 12.4, es decir

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

Vea el Ejemplo 12.7, donde se indica el código en lenguaje Python para calcular los valores de $\Phi(x)$. ■

Teorema 1.

Toda función de distribución $F(x)$ cumple las siguientes propiedades:

1. $F(x)$ es no decreciente.
2. $F(x)$ es continua a la derecha.
3. $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$.

Demostración:

1. Sean dos números reales x_1 y x_2 tales que $x_1 < x_2$. Como $(X \leq x_1) \subset (X \leq x_2)$, se tiene que $F(x_1) \leq F(x_2)$.

2. Sean c un número real, y $\{x_n\}$ una sucesión numérica tal que $x_n \downarrow 0$. Observe que $(X \leq c) = \bigcap_{n=1}^{\infty} (X \leq c + x_n)$. Entonces, utilizando el teorema de la continuidad de la probabilidad, obtenemos que $F(c) = \lim_n F(c + x_n) = F(c^+)$.

3. Se deduce inmediatamente utilizando el teorema de la continuidad de la probabilidad, teniendo en cuenta que si $x_n \uparrow +\infty$, entonces

$$\Omega = \bigcup_{n=1}^{\infty} (X \leq x_n) \quad \text{y} \quad \emptyset = \bigcap_{n=1}^{\infty} (X \leq -x_n).$$

Ejercicio 1.

Sea X una variable aleatoria no negativa que tiene densidad de probabilidad. Denotemos por $F(x)$ su función de distribución. Pruebe que

$$EX = \int_0^{\infty} (1 - F(x)) dx.$$

Observación 9 La fórmula del ejercicio anterior es válida para cualquier variable aleatoria no negativa, y no solo para variables aleatorias con densidad de probabilidad.

Teorema 2.

Sea $F(x)$ una función que cumple las tres propiedades de una función de distribución establecidas en el Teorema 1. Entonces existen un espacio de probabilidad, y una variable aleatoria definida sobre este cuya función de distribución es $F(x)$.

Demostración:

Tomemos el espacio de probabilidad $((0, 1), \mathcal{B}((0, 1)), \lambda)$. Definamos la *función inversa generalizada* de $F(x)$ de la siguiente manera,

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \text{ para todo } 0 < u < 1.$$

Cuando $F(x)$ es estrictamente creciente, la función inversa de $F(x)$ coincide con su función inversa generalizada. Realizaremos la demostración solo en el caso en que $F(x)$ es estrictamente creciente. En el caso general la prueba es laboriosa y no se realizará.

Sea la variable aleatoria $X = F^{-1}(U)$, donde U es una variable aleatoria con distribución uniforme, definida sobre el espacio de probabilidad $((0, 1), \mathcal{B}((0, 1)), \lambda)$. Luego X también está definida sobre $((0, 1), \mathcal{B}((0, 1)), \lambda)$. Entonces

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x). \quad (13.2)$$



El teorema anterior justifica que en muchas aplicaciones de la teoría de las probabilidades se trabaje con las funciones de distribución de las variables aleatorias, sin precisar cómo las variables están definidas; lo mismo sucede con las funciones de probabilidad y las densidades de probabilidad.

Note que toda función $f(x)$ no negativa, integrable, y con área bajo la curva de $f(x)$ igual a 1, es la densidad de probabilidad de cierta variable aleatoria. Esto es cierto porque $f(x)$, con las propiedades mencionadas, define una función de distribución.

Según la definición de función de distribución, queda claro que la distribución de probabilidad P_X de una variable aleatoria X determina su función de distribución $F(x)$. Puede demostrarse que, inversamente, la función de distribución de X determina a P_X , pero esta demostración requiere en el caso general de una teoría avanzada. En el caso de una variable aleatoria discreta la demostración es inmediata, puesto que los valores a_i de la variable aleatoria son los puntos de discontinuidad de $F(x)$, y las correspondientes probabilidades p_i son las longitudes de los saltos de $F(x)$ en estos puntos de discontinuidad, es decir, $p_i = F(a_i) - F(a_i^-)$. Especificada la función de probabilidad de una variable aleatoria X , es conocido que esta determina la distribución de probabilidad P_X (vea la Sección 10.4).

La función de distribución $F(x)$ y la distribución de probabilidad P_X son formas generales equivalentes de expresar la distribución de una variable aleatoria X . Dos formas particulares de expresar una distribución son, la densidad de probabilidad en el caso en que esta exista, y la función de probabilidad en el caso de una variable aleatoria discreta.

Puesto que la función de distribución $F(x)$ y la distribución de probabilidad P_X de una variable aleatoria X se determinan mutuamente, entonces en lugar de escribir

$$Eg(X) = \int_{\mathbb{R}} g(x) dP_X(x),$$

muchas veces se escribe

$$Eg(X) = \int_{\mathbb{R}} g(x) dF(x).$$

La igualdad (13.2) en la demostración del Teorema 2 prueba, además, la siguiente proposición.

Proposición 3.

Sean X una variable aleatoria con función de distribución $F(x)$ y $U \sim U(0, 1)$. Entonces la variable aleatoria $Y = F^{-1}(U)$ tiene igual función de distribución $F(x)$ que X .

Esta última proposición es de gran importancia práctica, puesto que en ella se establece el **método estándar de simulación** de variables aleatorias.

Ejercicio 2.

Pruebe que $U \sim U(0, 1)$ si, y solo si, $(1 - U) \sim U(0, 1)$.

Ejemplo 13.2 (Vea el Ejemplo 13.1). Sea $U \sim U(0, 1)$.

a) Si $X \sim U(a, b)$, entonces X se puede simular usando la variable aleatoria Y definida como

$$Y = F^{-1}(U) = (b - a)U + a.$$

b) Si $X \sim \text{Exp}(\lambda)$, entonces X se puede simular por la variable aleatoria Y_1 , definida por la fórmula

$$Y_1 = F^{-1}(U) = -\frac{\ln(1 - U)}{\lambda}.$$

Pero del Ejercicio 2 se deduce que X también puede simularse por la variable aleatoria Y_2 , definida como

$$Y_2 = -\frac{\ln(U)}{\lambda},$$

lo cual es más eficiente. ■

Como la función inversa de la función de distribución de una variable aleatoria con distribución normal no se puede expresar por medio de una fórmula exacta de manera explícita, entonces en este caso se usan técnicas especiales para la simulación. Con el siguiente código en lenguaje Python se generan n valores de una variable aleatoria $X \sim N(a, b^2)$.

```
from scipy.stats import norm
x=norm.rvs(a,b,n)
```

Ejercicio 3.

Sea $F(x)$ la función de distribución de una variable aleatoria X . Pruebe que entonces se cumplen las siguientes propiedades:

1. $P(a < X \leq b) = F(b) - F(a)$, si $a < b$
2. $P(X = c) = F(c) - F(c^-)$
3. $P(X < x) = F(x^-)$.

De la propiedad 2 del ejercicio anterior se deduce que la probabilidad de que una variable aleatoria tome el valor de un punto de continuidad de su función de distribución es igual a 0. Por lo tanto, si la función de distribución es continua (como es el caso de distribuciones con densidad de probabilidad), la probabilidad de que la variable aleatoria tome un valor particular es igual a 0.

Ejercicio 4.

El conjunto de puntos de discontinuidad de una función de distribución es finito o infinito numerable.

13.2. Variable aleatoria mixta

Un tipo de variable aleatoria que no es ni discreta ni tiene densidad de probabilidad, y es de interés tanto práctico como teórico, es la llamada *variable aleatoria mixta*. En este caso la función de distribución de la variable aleatoria es discontinua en al menos un punto, y estrictamente creciente y continua en algún subintervalo. A continuación se presentan dos ejemplos que ilustran este tipo de variable aleatoria.

Consideremos que la variable aleatoria X , con función de distribución $F(x)$, tiene densidad de probabilidad. Supongamos que X representa el tiempo de vida de un equipo, y que se planifica la sustitución preventiva del equipo al cabo de un tiempo T desde el comienzo de su funcionamiento. Sea $0 < F(t) < 1$, para $t > 0$. Entonces la variable aleatoria $Y = \min(X, T)$ representa el tiempo de trabajo efectivo del equipo.

Note que

$$P(Y > t) = \begin{cases} 1 - F(t), & \text{si } t < T \\ 0, & \text{si } t \geq T. \end{cases}$$

Luego la función de distribución de Y es continua excepto en el punto T . Esta variable aleatoria no es discreta, y tampoco tiene densidad de probabilidad; es una variable aleatoria mixta. De acuerdo con la Observación 9 se tiene que

$$EY = \int_0^T (1 - F(x))dx.$$

Consideremos ahora un dispositivo al cual llegan clientes para ser servidos. Asumamos que el tiempo de servicio y los tiempos entre llegadas de los clientes son variables aleatorias con densidad de probabilidad. Sea X el tiempo de espera por el comienzo del servicio de un cliente dado. Entonces X es igual a 0 si a la llegada del cliente el servidor está desocupado, y si está ocupado, X es igual al tiempo que transcurre hasta que son servidos todos los clientes que se encontraban en el sistema en el instante de su llegada. En general $P(X = 0) > 0$ y, por consiguiente, la función de distribución del tiempo de espera del cliente tiene un salto en el origen de coordenadas igual a $P(X = 0)$; puede probarse que para valores mayores que cero es una función continua.

13.3. Sobre estadística

Histograma de frecuencias

Sean $F(x)$ y $f(x)$ la función de distribución y la densidad de probabilidad, respectivamente, de una variable aleatoria X . Supongamos que $f(x)$ es continua. Entonces

$$P\left(t - \frac{h}{2} < X < t + \frac{h}{2}\right) = F\left(t + \frac{h}{2}\right) - F\left(t - \frac{h}{2}\right) = f(t)h + o(h^2).$$

Es decir, la probabilidad de que la variable aleatoria X tome valores en el intervalo $(t - \frac{h}{2}, t + \frac{h}{2})$, cuando h es pequeño, se calcula aproximadamente por la fórmula,

$$P\left(t - \frac{h}{2} < X < t + \frac{h}{2}\right) \approx f(t)h.$$

Por otra parte, si podemos observar un número grande n de valores de la variable aleatoria X (*muestra de tamaño n de valores observados de X*),

entonces utilizando el método estadístico de cálculo de una probabilidad se obtiene

$$P\left(t - \frac{h}{2} < X < t + \frac{h}{2}\right) \approx \frac{m}{n},$$

donde m es el número de valores de la muestra que pertenecen al intervalo $(t - \frac{h}{2}, t + \frac{h}{2})$. Combinando las dos últimas fórmulas aproximadas se halla que

$$f(t) \approx \frac{m}{nh}, \quad t \in \left(t - \frac{h}{2}, t + \frac{h}{2}\right).$$

Luego podemos determinar de forma aproximada la densidad de probabilidad correspondiente a una variable aleatoria X de la siguiente manera: observamos una muestra de tamaño grande n de valores de X , digamos, x_1, x_2, \dots, x_n ; dividimos el rango de valores de X en k subintervalos $[t_i, t_{i+1})$, donde $t_{i+1} = t_i + h$, $h > 0$, $i = 1, 2, \dots, k$; y contamos los valores m_i observados de X que pertenecen al i -ésimo intervalo. La siguiente función $\hat{f}(t)$, llamada **histograma de frecuencias** relativas, tendrá la forma aproximada de la densidad de probabilidad $f(x)$,

$$\hat{f}(t) = \frac{m_i}{nh}, \quad t \in [t_i, t_{i+1}), \quad i = 1, 2, \dots, k.$$

Estimación de parámetros

Anteriormente, en la Sección 9.3, se examinó cómo estimar una probabilidad y el valor esperado, ahora se continúa este examen.

Sea X_1, X_2, \dots, X_n una muestra aleatoria de la variable aleatoria X . Esto quiere decir que las variables aleatorias X_1, X_2, \dots, X_n son independientes e igualmente distribuidas que la variable aleatoria X . Denotemos $EX = a$ y $V(X) = \sigma_X^2$. Sea

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Entonces

$$E\bar{X}_n = a.$$

Además,

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_X^2 = \frac{\sigma_X^2}{n}.$$

De este último resultado se deduce que cuando n crece, la varianza de la media aritmética decrece.

Bajo las suposiciones hechas, los dos resultados obtenidos apuntan hacia la siguiente idea. Cuando el tamaño de la muestra n es grande, la media aritmética de observaciones de una variable aleatoria X constituye una variable aleatoria cuyos valores deben proporcionar buenas estimaciones del valor esperado de X .

Aunque posteriormente será mejor fundamentado, de las consideraciones anteriores parece claro que cuando el tamaño de la muestra n es grande, el valor esperado de una variable aleatoria X puede calcularse aproximadamente usando la media aritmética de los valores observados que constituyen la muestra. Es decir, si x_1, x_2, \dots, x_n , son los valores correspondientes a una muestra de X , entonces

$$EX \approx \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Esta fórmula aproximada sugiere cómo calcular aproximadamente los valores de los parámetros de algunas distribuciones. Por ejemplo, en el caso de la distribución normal, para el parámetro μ se tiene

$$\mu \approx \bar{x}.$$

Cuando la distribución de X es exponencial, el valor del parámetro λ se calcula aproximadamente de la siguiente manera,

$$EX = \frac{1}{\lambda} \approx \bar{x},$$

luego

$$\lambda \approx \frac{n}{\sum_{i=1}^n x_i}.$$

De manera similar, la varianza de una variable aleatoria X puede calcularse aproximadamente a partir de n valores observados, x_1, x_2, \dots, x_n , de X , por la fórmula

$$V(X) \approx \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

cuando n es grande, donde el miembro derecho de la expresión anterior es la desviación estándar de la muestra. En el caso de la distribución normal, se puede usar la fórmula aproximada anterior para estimar el parámetro σ^2 .

El siguiente código en lenguaje Python genera 1000 valores de una variable aleatoria $X \sim N(6, 4)$, dibuja el histograma de frecuencias (absolutas)

correspondiente a los datos generados utilizando 15 subintervalos, y calcula la media y la desviación estándar de estos datos.

```
from scipy.stats import norm
import matplotlib.pyplot as plt
import numpy as np
x = norm.rvs(6, 2, size=1000)
plt.hist(x, bins = 15)
media=np.mean(x)
desv_standar=np.std(x)
```

Finalmente, resumamos la idea general de un procedimiento estadístico que permite asumir, o no, determinada densidad de probabilidad. A partir de una muestra de la variable aleatoria de interés se construye el histograma de frecuencias correspondiente y, de acuerdo a la forma de este, se hace la hipótesis sobre un posible tipo de distribución de probabilidad (densidad). Con los datos de la muestra se estima el parámetro o los parámetros de los cuales depende la densidad de probabilidad. Por último, se utiliza un método estadístico (que no explicaremos en este texto) que permita decidir si, con una probabilidad de error pequeña dada, no hay concordancia entre la distribución de probabilidad que refleja la muestra y la hipótesis hecha; o en caso contrario, si los datos no muestran falta de concordancia con la hipótesis.

Muchas veces en las aplicaciones, la hipótesis sobre un tipo dado de densidad de probabilidad se realiza no sobre la base de un histograma de frecuencias, sino a partir de consideraciones teóricas o de la experiencia práctica. Para una variable aleatoria discreta el concepto de histograma de frecuencias no es válido; sin embargo se puede llevar a cabo una idea similar a la descrita en el párrafo anterior.

13.4. Método de Monte Carlo

En la Sección 2.3 se examinó un método de cálculo aproximado de integrales por medio de la generación de números aleatorios. A continuación presentamos un segundo método para el cálculo aproximado de integrales de funciones reales de una variable real. En general este método es más eficiente que el primero. El método es análogo al expuesto en el Ejemplo 9.6.

Sea la función $g(x)$ integrable en el intervalo $[a, b]$. Supongamos que se desea calcular la integral

$$\int_a^b g(x)dx.$$

Note que si X es una variable aleatoria con distribución uniforme en el intervalo (a, b) , entonces

$$E(g(X)) = \frac{1}{b-a} \int_a^b g(x) dx.$$

Por otro lado,

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

donde x_1, x_2, \dots, x_n son valores generados de la variable aleatoria X . Luego, para n grande, se tiene que

$$\int_a^b g(x) dx \approx \frac{b-a}{n} \sum_{i=1}^n g(x_i). \quad (13.3)$$

13.5. Funciones de una variable aleatoria

A menudo se conoce la distribución de una variable aleatoria X y se desea determinar la distribución de una variable aleatoria $Y = g(X)$, donde $g(x)$ es una función Boreliana. Esto no es difícil cuando X es una variable discreta (vea el Ejemplo 12.8). Casos de mayor interés son aquellos en los cuales la variable aleatoria X tiene densidad de probabilidad.

Ejemplo 13.3 Sea X una variable aleatoria con densidad de probabilidad continua $f_X(x)$. Hallemos la densidad de probabilidad $f_Y(y)$ de la variable aleatoria $Y = X^2$.

Denotemos por $F_X(x)$ y $F_Y(y)$ las funciones de distribución de las variables aleatorias X y Y , respectivamente.

Para $y \leq 0$ se tiene, $F_Y(y) = P(Y \leq y) = 0$. Sea ahora $y > 0$. Entonces

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Derivando $F_Y(y)$ se obtiene

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})].$$

■

Ejemplo 13.4 Supongamos que $X \sim N(0, 1)$. Hallemos la densidad de probabilidad de $Y = X^2$. Utilizando el resultado del ejemplo anterior, y teniendo en cuenta que

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

hallamos,

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}.$$

■

La densidad de probabilidad $f_Y(y)$ obtenida en el último ejemplo, corresponde a la distribución χ^2 con parámetro 1, o sea $\chi^2(1)$ (vea la Sección 12.5).

Teorema 4.

Sean X una variable aleatoria con densidad de probabilidad continua $f_X(x)$ y $g(x)$ una función estrictamente monótona y con derivada continua. Entonces la variable aleatoria $Y = g(X)$ tiene densidad de probabilidad dada por

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|,$$

donde $g^{-1}(y)$ es la función inversa de $g(x)$.

Demostración:

La demostración se realizará para el caso en que $g(x)$ es una función estrictamente creciente. El caso en que $g(x)$ es estrictamente decreciente se prueba de forma similar. Note que

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Derivando ahora $F_Y(y)$ se obtiene el resultado requerido. ■

Del teorema anterior se deduce, en particular, que si una variable aleatoria X tiene densidad de probabilidad continua $f_X(x)$ y $Y = aX + b$, donde a y b son números reales, con $a \neq 0$, entonces

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

Ejemplo 13.5 Sean $X \sim N(\mu, \sigma^2)$ y $Y = aX + b$, donde a y b son constantes reales, $a \neq 0$. Utilizando la fórmula anterior se comprueba fácilmente que $Y \sim N(a\mu + b, a^2\sigma^2)$. De manera similar se obtiene que $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. ■

13.6. Problemas y ejercicios

1. Si X una variable aleatoria con valores en \mathbb{R} , entonces $P(|X| < \infty) = 1$. Suponga ahora que X toma valores en \mathbb{R} y tiene función de distribución $F(x)$, pruebe que $P(X < +\infty) = \lim_{x \rightarrow +\infty} F(x)$.
2. Sea $F(x)$ la función de distribución de una variable aleatoria X . Suponga que $F(x)$ es estrictamente creciente y continua. Halle la distribución de la variable aleatoria $F(X)$.
3. *Suponga que usted tiene una cita. Sean c_1 el costo por unidad de tiempo de llegada tarde a la cita, y c_2 el costo por unidad de tiempo de llegada con antelación a la cita. Suponga además que el tiempo del viaje suyo hasta el lugar de la cita es una variable aleatoria con función de distribución diferenciable $F(x)$. Pruebe que el tiempo t de antelación a la cita, con el cual usted debe partir para minimizar el costo esperado, se determina por la ecuación,

$$F(t) = c_1 / (c_1 + c_2).$$

4. Una variable aleatoria X se dice que tiene *distribución de Laplace* con parámetro λ si su densidad de probabilidad se expresa, $f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$, $x \in \mathbb{R}$, $\lambda > 0$,
 - a) Halle la esperanza matemática de X .
 - b) Sean $X \sim \text{Exp}(\lambda)$ y $U \sim U(0, 1)$ variables aleatorias independientes. Compruebe que la variable aleatoria Y definida como, $Y = X$ si $U \leq 1/2$, y $Y = -X$ si $U > 1/2$, tiene distribución de Laplace.
 - c) Proponga un método para simular una variable aleatoria con distribución de Laplace.
5. Sea $X \sim \text{Exp}(\lambda)$. Implemente un programa en lenguaje Python para generar una muestra de valores de X por medio de la simulación. Construya el histograma de frecuencias correspondiente a la muestra y compare la media aritmética con EX . Tome $\lambda = 0,01$.

6. Sea X una variable aleatoria con densidad de probabilidad, que tiene función de distribución $F(x)$. Pruebe que si $E|X| < \infty$, entonces

$$EX = \int_0^{\infty} (1 - F(x))dx - \int_{-\infty}^0 F(x)dx.$$

7. *Suponga que usted llega en un instante aleatorio a una parada de ómnibus donde puede tomar una de dos rutas. Los ómnibus de cada una de las rutas llegan uno tras otro a intervalos de diez minutos, y de manera independiente los de rutas diferentes. ¿Cuál es el valor esperado del tiempo de espera suyo? **R:** 10/3.
8. Sean $X \sim U(a, b)$ y $Z = \min(X, T)$, donde $a > 0$ y T es un número real, $T > 0$. Analice cuándo Z tiene densidad de probabilidad y cuándo no. Calcule EZ .
9. Formule cómo calcular aproximadamente, usando técnicas de simulación, la integral de una función $g(x)$ en el intervalo infinito $(0, \infty)$. Calcule aproximadamente, y de forma exacta por métodos analíticos, las siguientes integrales:

$$\int_0^2 x(1 - x/2) dx \qquad \int_0^{\infty} e^{-x^2} dx.$$

Compare el resultado con el valor exacto.

Sugerencia: Vea la fórmula 13.3.

10. Sea $Z \sim N(0, 1)$. Halle la densidad de probabilidad de la variable aleatoria $Y = |Z|$.
11. Sea X con distribución uniforme en el intervalo $(-\pi/2, \pi/2)$. Halle la densidad de probabilidad de $Y = \tan(X)$. Comente sobre el valor esperado de Y .
12. Se dice que una variable aleatoria X tiene *distribución lognormal* con parámetros μ y σ^2 si $Y = \ln X$ tiene distribución $N(\mu, \sigma^2)$. Halle la densidad de probabilidad de X . Compruebe que $EX = e^{\mu + \sigma^2/2}$.

COMPLEMENTOS

14.1. Identidad de Wald

El siguiente teorema, referido al valor esperado de una suma aleatoria, tiene gran importancia en el estudio de diversos modelos estocásticos.

Teorema 1 (A.Wald).

Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e igualmente distribuidas, y sea N una variable aleatoria con valores enteros no negativos tal que el evento $(N = n)$ es independiente de las variables aleatorias X_{n+1}, X_{n+2}, \dots , para todo $n \geq 1$. Supongamos que X_1 es no negativa, o en el caso general, que $E|X_1| < \infty$, y además que $EN < \infty$. Entonces

$$E \left(\sum_{i=1}^N X_i \right) = EX_1 \cdot EN.$$

Demostración:

Tienen lugar las igualdades,

$$\begin{aligned}
 E\left(\sum_{n=1}^N X_n\right) &= E\left(\sum_{n=1}^{\infty} X_n I_{(N \geq n)}\right) = \sum_{n=1}^{\infty} E(X_n I_{(N \geq n)}) \\
 &= \sum_{n=1}^{\infty} EX_n \cdot EI_{(N \geq n)} = EX_1 \sum_{n=1}^{\infty} EI_{(N \geq n)} \\
 &= EX_1 \sum_{n=1}^{\infty} P(N \geq n) = EX_1 \cdot EN.
 \end{aligned}$$

Cuando las variables aleatorias de la sucesión $\{X_n\}$ son no negativas, el intercambio entre los símbolos de valor esperado y suma en la segunda igualdad se justifica por el corolario del TCM, y en el caso general se justifica utilizando el teorema de Fubini (de la teoría de la medida). Por otra parte, note que

$$(N \geq n) = (N \leq n-1)^c = \left(\bigcup_{k=1}^{n-1} (N = k)\right)^c,$$

y que por las suposiciones hechas, cada uno de los sucesos $(N = 0), (N = 1), \dots, (N = n-1)$ mutuamente excluyentes, y cada una de las variables aleatorias X_n, X_{n+1}, \dots , son independientes. Luego, en particular,

$$\bigcup_{k=1}^{n-1} (N = k) \text{ y } X_n, \text{ para todo } n \geq 2,$$

son independientes y, por lo tanto, también lo son $I_{(N \geq n)}$ y $X_n, n \geq 1$, de lo cual se deduce el cumplimiento de la tercera igualdad. ■

Es importante resaltar, que el teorema anterior se cumple en particular, cuando la variable aleatoria N es independiente de todas las variables aleatorias X_1, X_2, \dots .

Como un ejemplo de aplicación de la identidad de Wald consideremos el siguiente problema.

Modelo de sustitución preventiva. El funcionamiento de cierto sistema se describe de la siguiente manera. Se dispone de componentes cuyos tiempos de vida son las variables aleatorias X_1, X_2, \dots , independientes e igualmente distribuidas, con función de distribución $F(t)$. En cada instante de

tiempo trabaja un único componente. En muchas situaciones prácticas los componentes envejecen y el fallo inesperado de estos puede tener consecuencias graves, luego es natural plantearse la sustitución preventiva de los componentes al cabo de un tiempo T después del inicio de su trabajo (no analizaremos aquí cómo determinar el valor de T). Entonces los componentes se sustituyen de forma consecutiva por otros nuevos, o porque fallaron o de manera preventiva al cabo del tiempo T de trabajo. Un problema de interés consiste en hallar el valor esperado del tiempo entre dos sustituciones consecutivas debido al fallo de los componentes.

El tiempo transcurrido entre dos sustituciones consecutivas (por fallo o por sustitución preventiva) lo denotaremos por Y y se expresa

$$Y = \min(X, T),$$

donde X es una variable con función de distribución $F(t)$. Asumiremos que $0 < F(t) < 1$, para todo $t > 0$. Recordemos que en la Sección 13.2 se consideró la variable aleatoria Y como un ejemplo de variable mixta. En esa sección se halló que

$$EY = \int_0^T (1 - F(x))dx.$$

Sea τ el tiempo entre dos sustituciones consecutivas debido al fallo de los componentes. Entonces

$$\tau = \sum_{k=1}^N Y_k,$$

donde las variables aleatorias $Y_k = \min(X_k, T)$, $k \geq 1$, son independientes y tienen igual distribución que Y , y N es una variable aleatoria que representa el número de sustituciones hasta que por primera vez la sustitución se debe al fallo inesperado de un componente. Puesto que

$$(N = n) = (X_1 > T, \dots, X_{n-1} > T, X_n < T), \quad n \geq 1.$$

el evento $(N = n)$ es independiente de las variables aleatorias Y_{n+1}, Y_{n+2}, \dots . Además, N tiene la distribución geométrica, $P(N = n) = (1 - p)^{n-1}p$, $n \geq 1$, donde $p = P(X < T)$. Entonces usando la identidad de Wald se obtiene

$$E\tau = EY \cdot EN = \frac{EY}{p} = \frac{\int_0^T (1 - F(x))dx}{F(T)}.$$

14.2. Espacio L^2 de variables aleatorias

Denotemos por L^1 el conjunto de todas las variables aleatorias (v.a) integrables, es decir,

$$L^1 = \{X \text{ v.a.} : E|X| < \infty\}.$$

Si $X, Y \in L^1$, y a y b son números reales, entonces $aX + bY \in L^1$, luego L^1 es un subespacio vectorial del espacio vectorial de todas las variables aleatorias.

Sean X y Y elementos de L^1 . Note que la relación $X = Y$ c.s constituye una relación de equivalencia, y como la suma de dos variables aleatorias y el producto de una variable aleatoria por un número real conservan la igualdad c.s, entonces el conjunto de las clases de equivalencia es un espacio vectorial. Por otra parte, las variables aleatorias pertenecientes a una misma clase tienen igual valor esperado.

Identifiquemos cada clase de equivalencia con un único elemento representante de la clase, por ejemplo, en el caso $X = 0$ c.s, consideraremos solo la variable aleatoria idénticamente igual a 0. En lugar de considerar el espacio vectorial de las clases de equivalencia, consideremos el conjunto constituido por los representantes de las clases, el cual, con un abuso de notación, también denotaremos por L^1 .

Sea ahora

$$L^2 = \{X \text{ v.a.} : EX^2 < \infty\}.$$

Denotaremos también por L^2 el conjunto de las variables aleatorias de cuadrado integrables que contiene únicamente los representantes de cada clase de equivalencia. Si $X \in L^2$, entonces $X \in L^1$. El conjunto L^2 es un subespacio vectorial de L^1 .

Sean $X, Y \in L^2$. Entonces, $XY \in L^1$. No es difícil comprobar que

$$\langle X, Y \rangle = EXY$$

es un producto escalar definido sobre L^2 . Este producto escalar induce la norma

$$\|X\| = \sqrt{EX^2}$$

y, por consiguiente, la distancia

$$d(X, Y) = \|X - Y\| = \sqrt{E(X - Y)^2}.$$

Consideremos las variables aleatorias $X_1 = X - EX$ y $Y_1 = Y - EY$. Asumiremos que $V(X) > 0$ y $V(Y) > 0$. El ángulo θ entre X_1 y Y_1 queda

definido por

$$\cos \theta = \frac{\langle X_1, Y_1 \rangle}{\|X_1\| \cdot \|Y_1\|} = \frac{E(X - EX)(Y - EY)}{\sqrt{V(X)}\sqrt{V(Y)}}.$$

El numerador de la fórmula anterior es la covarianza de X y Y , y $\cos \theta$ es el coeficiente de correlación ρ .

De la Proposición 7 se conoce que $-1 \leq \rho \leq 1$. Cuando $\rho = 0$ las variables X_1 y Y_1 son ortogonales, y las variables X y Y son no correlacionadas.

Dadas las variables aleatorias linealmente independientes, $X_1, \dots, X_n \in L^2$, consideremos ahora el siguiente subespacio vectorial de L^2 ,

$$L = \left\{ \sum_{i=1}^n \alpha_i X_i : \alpha_i \in \mathbb{R}, i = 1, 2, \dots, n \right\}.$$

Sea $Y \in L^2$. Busquemos $\hat{Y} \in L$ que mejor aproxima a Y en el siguiente sentido:

$$d(Y, \hat{Y}) = \|Y - \hat{Y}\| = \min_{Z \in L} \|Y - Z\|. \quad (14.1)$$

Como es conocido, \hat{Y} es la proyección de Y sobre L y satisface (14.1) si, y solo si,

$$\langle Y - \hat{Y}, X_i \rangle = 0, \quad i = 1, 2, \dots, n. \quad (14.2)$$

Note que (14.2), donde $\hat{Y} = \sum_{j=1}^n \alpha_j X_j$, puede escribirse de la siguiente manera,

$$\langle Y, X_i \rangle = \sum_j \alpha_j \langle X_j, X_i \rangle, \quad i = 1, 2, \dots, n. \quad (14.3)$$

Examinaremos dos casos particulares:

1. Sean $n = 1$ y $X_1 = 1$. En este caso $L = \{\alpha_1 \cdot 1 : \alpha_1 \in \mathbb{R}\}$ y (14.3) toma la forma

$$\langle Y, 1 \rangle = \langle \alpha_1, 1 \rangle,$$

de donde, $\alpha_1 = EY$. Luego, $\hat{Y} = EY$ y

$$d(Y, \hat{Y}) = \|Y - \hat{Y}\| = \|Y - EY\| = \sqrt{E(Y - EY)^2} = \sigma_Y.$$

2. Sean $n = 2$, $X_1 \equiv 1$ y $X_2 = X \in L^2$, $X \neq X_1$. Ahora,

$$L = \{\alpha_1 X + \alpha_2 : \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

De (14.3) se tiene el siguiente sistema de ecuaciones, del cual debemos hallar α_1 y α_2 :

$$\begin{aligned} EY &= \langle Y, 1 \rangle = \langle \alpha_1 X + \alpha_2, 1 \rangle = \alpha_1 EX + \alpha_2, \\ EXY &= \langle Y, X \rangle = \langle \alpha_1 X + \alpha_2, X \rangle = \alpha_1 EX^2 + \alpha_2 EX. \end{aligned} \quad (14.4)$$

De la primera ecuación del sistema se tiene, $\alpha_2 = EY - \alpha_1 EX$, y sustituyendo en la ecuación

$$\hat{Y} = \alpha_1 X + \alpha_2,$$

se halla que

$$\hat{Y} = \alpha_1(X - EX) + EY. \quad (14.5)$$

Sustituyendo ahora la expresión de α_2 en la segunda ecuación del sistema (14.4) y despejando, se obtiene

$$\alpha_1 = \frac{EXY - EXEY}{V(X)} = \rho \frac{\sigma_Y}{\sigma_X}.$$

Finalmente, sustituyendo este valor de α_1 en (14.5) se obtiene

$$\hat{Y} = EY + \rho \frac{\sigma_Y}{\sigma_X}(X - EX).$$

Esta última expresión se denomina **recta de regresión de Y sobre X** .

El valor de $d(Y, \hat{Y})$ se halla,

$$d^2(Y, \hat{Y}) = E(Y - \hat{Y})^2 = \sigma_Y^2(1 - \rho^2),$$

de donde

$$d(Y, \hat{Y}) = \sigma_Y \sqrt{1 - \rho^2}.$$

En el Ejercicio 6 de la Sección 12.6 se estableció, que si entre X y Y existe una dependencia lineal c.s, entonces $\rho^2 = 1$. El siguiente teorema establece el recíproco de esta afirmación.

Teorema 2.

Si $\rho = 1$ o $\rho = -1$, entonces Y depende linealmente de X casi seguramente.

Demostración:

Notemos que si $\rho = 1$ o $\rho = -1$, entonces

$$d(Y, \hat{Y}) = E(Y - \hat{Y})^2 = 0.$$

Pero como $(Y - \hat{Y})^2 \geq 0$, esto significa que $(Y - \hat{Y})^2 = 0$ c.s, es decir, $Y = \hat{Y}$ c.s. Puesto que \hat{Y} depende linealmente de X , entonces Y depende linealmente de X c.s. ■

Parte III

Vectores aleatorios

VECTORES ALEATORIOS

En capítulos anteriores se consideró la distribución de una única variable aleatoria, en este capítulo se introduce la noción de distribución conjunta de varias variables aleatorias y se tratan diversos resultados asociados a esta distribución.

La mayoría de las definiciones y resultados se expondrán para dos variables aleatorias, pero estos se extienden de manera natural para un número arbitrario finito de variables aleatorias.

15.1. Vectores aleatorios

En la definición de un vector aleatorio interviene la σ -álgebra de los Borelianos $\mathcal{B}(\mathbb{R}^n)$, $n \geq 2$; recordemos que esta σ -álgebra fue definida en la Sección 10.2.

Si X y Y son variables aleatorias definidas en un mismo espacio de probabilidad (Ω, \mathcal{F}, P) , podría interesar calcular la probabilidad de que X tome valores en B_1 y Y en B_2 , donde B_1 y B_2 son Borelianos de \mathbb{R} , es decir, la probabilidad

$$P(\{\omega \in \Omega : \omega \in X^{-1}(B_1) \cap Y^{-1}(B_2)\}) = P((X \in B_1) \cap (Y \in B_2)),$$

que de manera abreviada se denota

$$P(X \in B_1, Y \in B_2).$$

Más general, puede ser de interés calcular la probabilidad $P((X, Y) \in B)$ de que el par (X, Y) tome valores en una región $B \in \mathcal{B}(\mathbb{R}^2)$.

Definición 1.

Sea el espacio medible (Ω, \mathcal{F}) . Una función $Z : \Omega \rightarrow \mathbb{R}^n, (\mathcal{F}, \mathcal{B}(\mathbb{R}^n))$ -medible, se denomina **vector aleatorio**, o variable aleatoria vectorial.

A veces se dice que el vector aleatorio Z está definido sobre el espacio medible (Ω, \mathcal{F}) , o que está definido sobre el correspondiente espacio de probabilidad (Ω, \mathcal{F}, P) considerado.

Análogamente al caso de una única variable aleatoria, la σ -álgebra $\sigma(Z)$, generada por el vector aleatorio Z , es la colección

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

Un vector aleatorio Z se representa por $Z = (X_1, X_2, \dots, X_n)$, donde $X_i, i = 1, 2, \dots, n$, son las funciones coordenadas de Z .

La siguiente definición extiende el concepto de función Boreliana.

Definición 2.

Una función $g : \mathbb{R}^n \rightarrow \mathbb{R}^m, n, m = 1, 2, \dots$, se dice que es una **función Boreliana**, si g es $(\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^m))$ -medible.

Si $Z = (X_1, X_2, \dots, X_n)$ es un vector aleatorio y $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una función Boreliana, entonces no es difícil ver que, análogamente al caso de una sola variable aleatoria, la función compuesta $g(Z)$ es una variable aleatoria cuando $m = 1$, y es un vector aleatorio cuando $m > 1$.

Sea $Z = (X, Y)$. El próximo teorema muestra que las coordenadas de Z son variables aleatorias, por eso también llamamos a Z **vector de variables aleatorias**.

Teorema 1.

Sean el espacio medible (Ω, \mathcal{F}) y $Z = (X, Y)$. Entonces la función $Z : \Omega \rightarrow \mathbb{R}^2$ es un vector aleatorio sobre (Ω, \mathcal{F}) si, y solo si, X y Y son variables aleatorias sobre (Ω, \mathcal{F}) .

Demostración:

Solo si. La función coordenada $g_1(x, y) = x$ es una función Boreliana, pues para todo Boreliano $B \in \mathcal{B}(\mathbb{R})$ se tiene que $g_1^{-1}(B) = B \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^2)$. De forma similar, $g_2(x, y) = y$ es una función Boreliana. Luego, $g_1(Z) = X$ y $g_2(Z) = Y$ son variables aleatorias.

Si. Sean $B_1, B_2 \in \mathcal{B}(\mathbb{R})$. Entonces

$$Z^{-1}(B_1 \times B_2) = \{\omega : X(\omega) \in B_1, Y(\omega) \in B_2\} = X^{-1}(B_1) \cap Y^{-1}(B_2) \in \mathcal{F},$$

y como $\{B_1 \times B_2 : B_1, B_2 \in \mathcal{B}(\mathbb{R})\}$ es un conjunto generador de la σ -álgebra $\mathcal{B}(\mathbb{R}^2)$, por el Teorema 3 de la Sección 10.6 queda probado que Z es un vector aleatorio. ■

15.2. Distribución de vectores aleatorios

En lo que sigue asumiremos que las variables aleatorias están definidas en el mismo espacio de probabilidad (Ω, \mathcal{F}, P) .

Las dos próximas definiciones, análogas a las correspondientes al caso de una única variable aleatoria, constituyen dos formas generales equivalentes de expresar la distribución de un vector de variables aleatorias.

Definición 3.

Dado un vector aleatorio (X, Y) , la medida de probabilidad $P_{(X,Y)}$, definida sobre $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ por la fórmula

$$P_{(X,Y)}(B) = P((X, Y)^{-1}(B)), \quad \text{para todo } B \in \mathcal{B}(\mathbb{R}^2),$$

se denomina **distribución de probabilidad** del vector aleatorio (X, Y) . A $P_{(X,Y)}$ también se le llama probabilidad inducida por el vector aleatorio (X, Y) .

Definición 4.

Se llama **función de distribución conjunta** de las variables aleatorias X y Y , o función de distribución del vector aleatorio (X, Y) , a la función $F(x, y)$ definida como

$$F(x, y) = P((X \leq x) \cap (Y \leq y)) = P(X \leq x, Y \leq y), \quad x, y \in \mathbb{R}.$$

Dos vectores aleatorios Z_1 y Z_2 se dicen independientes, si las σ -álgebras $\sigma(Z_1)$ y $\sigma(Z_2)$ son independientes.

En el siguiente teorema se enuncian las cuatro propiedades que caracterizan la función de distribución de un vector aleatorio de dos variables aleatorias.

Teorema 2.

La función de distribución $F(x, y)$ del vector aleatorio (X, Y) cumple las siguientes propiedades:

1. $F(x, y)$ es una función no decreciente con respecto a cada variable.
2. $F(x, y)$ es continua a la derecha con respecto a cada variable.
3. $F(+\infty, +\infty) = \lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1.$
 $F(-\infty, y) = \lim_{x \rightarrow -\infty} F(x, y) = 0.$
 $F(x, -\infty) = \lim_{y \rightarrow -\infty} F(x, y) = 0.$
4. $F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0$, para todos a, b, c y d tales que $a \leq b$ y $c \leq d$.

Las tres primeras propiedades del teorema anterior, y sus demostraciones, son análogas a las tres propiedades que caracterizan la función de distribución de una variable aleatoria. La cuarta propiedad es peculiar de la función de distribución de vectores aleatorios y se obtiene verificando que

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c), \quad (15.1)$$

para todos a, b, c y d tales que $a \leq b$ y $c \leq d$.

Aunque no será demostrado, se cumple que cualquier función $F(x, y)$ que satisfaga las cuatro propiedades del Teorema 2 es función de distribución de cierto vector de variables aleatorias (X, Y) .

El siguiente ejemplo ilustra cómo una función $F(x, y)$ puede cumplir las tres primeras propiedades del teorema anterior, y no cumplir la cuarta: Sea $F(x, y) = 0$ si $x + y < 1$, y $F(x, y) = 1$ si $x + y \geq 1$. Es fácil ver que esta función cumple las tres primeras propiedades del Teorema 2; pero la cuarta propiedad no se cumple, puesto que $F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = -1$.

Ejercicio 1.

Pruebe las primeras tres propiedades de una función de distribución conjunta enunciadas en el Teorema 2.

El siguiente teorema muestra cómo, dada la función de distribución conjunta $F(x, y)$ del vector aleatorio (X, Y) , se obtienen las funciones de distribución $F_1(x)$ y $F_2(y)$ de X y Y , respectivamente. A estas distribuciones se les llama **distribuciones marginales**.

Teorema 3.

Sean $F(x, y)$, $F_1(x)$ y $F_2(y)$ las funciones de distribución del vector aleatorio (X, Y) y de las variables aleatorias X y Y , respectivamente. Entonces

$$F_1(x) = P(X \leq x) = \lim_{y \rightarrow +\infty} F(x, y) = F(x, +\infty),$$

$$F_2(y) = P(Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y) = F(+\infty, y).$$

Demostración:

Sea una sucesión numérica $\{y_n\}$ tal que $y_n \uparrow +\infty$. Entonces usando la propiedad de continuidad de la probabilidad se obtiene,

$$\begin{aligned} \lim_{y \rightarrow \infty} F(x, y) &= \lim_n P((X \leq x) \cap (Y \leq y_n)) = P\left(\bigcup_{n=1}^{\infty} (X \leq x) \cap (Y \leq y_n)\right) \\ &= P\left((X \leq x) \cap \bigcup_{n=1}^{\infty} (Y \leq y_n)\right) = P(X \leq x) = F_1(x). \end{aligned}$$

■

De las distribuciones marginales de X y Y en general no es posible obtener la distribución conjunta de (X, Y) . El siguiente teorema establece que cuando X y Y son independientes esto sí es posible.

Teorema 4.

Sean $F(x, y)$, $F_1(x)$ y $F_2(y)$ las funciones de distribución del vector aleatorio (X, Y) y de las variables aleatorias X y Y , respectivamente. Una condición necesaria y suficiente para que las variables aleatorias X y Y sean independientes es que se cumpla la igualdad

$$F(x, y) = F_1(x)F_2(y), \text{ para todos } x, y \in \mathbb{R}.$$

En el teorema anterior, la validez de la necesidad se ve de manera inmediata, puesto que si $\sigma(X)$ y $\sigma(Y)$ son independientes, entonces los eventos $(X \leq x)$ y $(Y \leq y)$ son independientes cualquiera que sean x y y números reales. La demostración de la suficiencia requiere una teoría avanzada. Pero al menos es posible probar el resultado que se enuncia en el siguiente ejercicio.

Ejercicio 2.

Sean $F(x, y)$, $F_1(x)$ y $F_2(y)$ las funciones de distribución de (X, Y) , X y Y , respectivamente. Pruebe que si $F(x, y) = F_1(x)F_2(y)$, para todos $x, y \in \mathbb{R}$, entonces se cumple que los eventos $(a < X \leq b)$ y $(c < Y \leq d)$ son independientes.

Sugerencia: Pruebe primero la fórmula 15.1.

15.3. Vector de variables aleatorias discretas

Sean el espacio de probabilidad (Ω, \mathcal{F}, P) , y las variables aleatorias discretas X y Y definidas en este espacio. Denotemos por a_1, a_2, \dots los valores de la variable aleatoria X , y por b_1, b_2, \dots los valores de la variable aleatoria Y .

En este caso, la *distribución conjunta* de las variables aleatorias X y Y también se determina por la **función de probabilidad conjunta** $f(x, y)$ definida sobre el conjunto de valores (a_i, b_j) , $i, j \geq 1$, del vector (X, Y) como

$$f(a_i, b_j) = P(X = a_i, Y = b_j), \quad i, j \geq 1.$$

Note que

$$\sum_{i,j=1}^{\infty} f(a_i, b_j) = 1.$$

Teorema 5.

Sea (X, Y) un vector de variables aleatorias discretas con función de probabilidad $f(a_i, b_j)$, $i, j \geq 1$. Entonces las funciones de probabilidad $f_1(a_i)$, $i \geq 1$, y $f_2(b_j)$, $j \geq 1$, de X y Y se determinan por las fórmulas,

$$f_1(a_i) = \sum_{j=1}^{\infty} f(a_i, b_j), \quad i \geq 1 \quad \text{y} \quad f_2(b_j) = \sum_{i=1}^{\infty} f(a_i, b_j), \quad j \geq 1.$$

Demostración:

La función de probabilidad $f_1(a_i)$, $i \geq 1$, de X se obtiene de la siguiente manera,

$$\sum_{j=1}^{\infty} f(a_i, b_j) = \sum_{j=1}^{\infty} P(X = a_i, Y = b_j) = P(X = a_i) = f_1(a_i), \quad i \geq 1.$$

De manera análoga se halla la función de probabilidad $f_2(b_j)$, $j \geq 1$, de Y . ■

A las funciones de probabilidad $f_1(x)$ y $f_2(y)$, obtenidas según las fórmulas del teorema anterior a partir de la función de probabilidad conjunta $f(x, y)$, se les llama **funciones de probabilidad marginales**.

Similar al caso de una sola variable aleatoria, si (X, Y) es un vector de variables aleatorias discretas con valores $\{a_i, b_j\}$, $i, j \geq 1$, y $g(x, y)$ una función Boreliana, se cumple que

$$E(g(X, Y)) = \sum_{i,j} g(a_i, b_j) P(X = a_i, Y = b_j),$$

si $g(x, y)$ es no negativa, o en el caso general, si la serie anterior converge absolutamente. En particular,

$$EXY = \sum_{i,j} a_i b_j P(X = a_i, Y = b_j).$$

15.4. Vector de variables aleatorias con densidad de probabilidad

Recordemos que en la Sección 10.5 se hizo notar que las variables aleatorias discretas X y Y son independientes si, y solo si,

$$P(X = a_i, Y = b_j) = P(X = a_i)P(Y = b_j), \quad i, j = 1, 2, \dots$$

Una distribución conjunta de variables aleatorias discretas de gran importancia es la distribución multinomial (vea la Sección 7.3). Se dice que el vector de variables aleatorias (X_1, X_2, \dots, X_r) , $r \geq 2$, tiene **distribución multinomial** con parámetros n y p_1, p_2, \dots, p_r , si

$$P(X_1 = k_1, X_2 = k_2, \dots, X_r = k_r) = \frac{n!}{k_1!k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r},$$

donde, n es un número natural, $p_i \geq 0$, $1 \leq i \leq r$, $\sum_{i=1}^r p_i = 1$ y k_1, k_2, \dots, k_r son números enteros no negativos tales que $\sum_{i=1}^r k_i = n$.

Puede demostrarse que para la distribución multinomial la distribución marginal de X_i es Binomial(n, p_i), $i = 1, 2, \dots, r$.

15.4. Vector de variables aleatorias con densidad de probabilidad

En lo adelante las integrales múltiples se considerarán en el sentido de Riemann.

Definición 5.

Se dice que el vector de variables aleatorias (X, Y) tiene **densidad de probabilidad**, si existe una función Boreliana no negativa $f(x, y)$, denominada densidad de probabilidad, tal que

$$P((X, Y) \in B) = \iint_B f(x, y) dx dy,$$

para todo $B \in \mathcal{B}(\mathbb{R}^2)$.

A la densidad de probabilidad $f(x, y)$ de un vector de variables aleatorias (X, Y) también se le llama *densidad conjunta* del vector de variables aleatorias,

y es una forma de expresar su **distribución conjunta**. Naturalmente se cumple que

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = 1.$$

De forma equivalente a la Definición 5, se puede definir un vector de variables aleatorias (X, Y) con densidad de probabilidad $f(x, y)$, como aquel cuya función de distribución $F(x, y)$ se expresa

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv, \quad (15.2)$$

donde $f(x, y)$ es la densidad de probabilidad del vector. En los puntos de continuidad de $f(x, y)$ se tiene,

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Ejemplo 15.1 Se dice que el vector aleatorio (X, Y) tiene **distribución uniforme** en la región D , $D \in \mathcal{B}(\mathbb{R}^2)$, $0 < \lambda(D) < \infty$ (λ -medida de Lebesgue en $\mathcal{B}(\mathbb{R}^2)$), lo cual se denota $(X, Y) \sim U(D)$, si este tiene densidad de probabilidad $f(x, y)$ constante en la región D , e igual a 0 fuera de esta región. De aquí se deduce que el valor constante de la densidad de probabilidad en la región D es igual a $1/\lambda(D)$. Por otro lado,

$$P((X, Y) \in B) = \iint_B f(x, y) dx dy = \frac{\lambda(B)}{\lambda(D)}, \quad \text{para todo } B \subset D, B \in \mathcal{B}(\mathbb{R}^2).$$

■

Ejemplo 15.2 Sea la densidad de probabilidad $f(x, y)$ de cierto vector aleatorio (X, Y) definida como $f(x, y) = \lambda \mu e^{-(\lambda x + \mu y)}$ para $x, y > 0$, y $f(x, y) = 0$ en otro caso, con $\lambda, \mu > 0$. Calculemos $P(X < Y)$.

$$\begin{aligned} P(X < Y) &= \iint_{x < y} \lambda \mu e^{-(\lambda x + \mu y)} dx dy = \int_0^{+\infty} \mu e^{-\mu y} \left(\int_0^y \lambda e^{-\lambda x} dx \right) dy \\ &= \mu \int_0^{+\infty} e^{-\mu y} (1 - e^{-\lambda y}) dy = 1 - \frac{\mu}{\mu + \lambda} = \frac{\lambda}{\mu + \lambda}. \end{aligned}$$

■

Teorema 6.

Sea el vector aleatorio (X, Y) con densidad de probabilidad $f(x, y)$. Entonces existen las densidades de probabilidad $f_1(x)$ y $f_2(y)$ de X y Y , y

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_2(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Demostración:

La función de distribución $F_1(x)$ de X se expresa

$$F_1(x) = F(x, +\infty) = \iint_{(-\infty, x] \times (-\infty, +\infty)} f(u, v) du dv = \int_{-\infty}^x \left(\int_{-\infty}^{+\infty} f(u, v) dv \right) du,$$

donde en la última igualdad se usó el teorema de Fubini. Luego,

$$f_1(u) = \int_{-\infty}^{+\infty} f(u, v) dv.$$

Con una teoría avanzada se demuestra que $f_1(x)$ es una función Boreliana. El resultado para $f_2(y)$ se obtiene de forma análoga. ■

A las densidades de probabilidad $f_1(x)$ y $f_2(y)$, obtenidas según las fórmulas del teorema anterior a partir de la densidad de probabilidad conjunta $f(x, y)$, se les llama **densidades de probabilidad marginales**.

Ejemplo 15.3 Consideremos la densidad de probabilidad $f(x, y)$ del ejemplo 15.2. Utilizando el Teorema 6, hallamos que la densidad de probabilidad $f_1(x)$ de X es igual a 0 si $x \leq 0$, y para $x > 0$ obtenemos,

$$f_1(x) = \int_0^{+\infty} \lambda \mu e^{-(\lambda x + \mu y)} dy = \lambda e^{-\lambda x}.$$

O sea, $X \sim \text{Exp}(\lambda)$. Similarmente se halla que $Y \sim \text{Exp}(\mu)$. ■

Teorema 7.

Sean X y Y variables aleatorias independientes con densidades de probabilidad respectivas $f_1(x)$ y $f_2(y)$. Entonces existe la densidad de probabilidad $f(x, y)$ del vector (X, Y) y

$$f(x, y) = f_1(x)f_2(y) \quad x, y \in \mathbb{R}.$$

Demostración:

Utilizando el Teorema 4 y el teorema de Fubini, se obtiene

$$\begin{aligned} F(x, y) &= F_1(x)F_2(y) = \int_{-\infty}^x f_1(u)du \int_{-\infty}^y f_2(v)dv \\ &= \int_{-\infty}^x \int_{-\infty}^y f_1(u)f_2(v)dudv. \end{aligned}$$

Por consiguiente, de acuerdo con la fórmula (15.2) $f_1(x)f_2(y)$ es la densidad de probabilidad de (X, Y) . ■

Teorema 8.

Sean $f(x, y)$, $f_1(x)$ y $f_2(y)$ las densidades de probabilidad de (X, Y) , X y Y , respectivamente. Si

$$f(x, y) = f_1(x)f_2(y), \quad x, y \in \mathbb{R}.$$

entonces las variables aleatorias X y Y son independientes.

Demostración:

Usando el teorema de Fubini se tiene,

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(u, v)dudv = \int_{-\infty}^x \int_{-\infty}^y f_1(u)f_2(v)dudv \\ &= \int_{-\infty}^x f_1(u)du \int_{-\infty}^y f_2(v)dv = F_1(x)F_2(y), \end{aligned}$$

de donde por el Teorema 4 se deduce el resultado requerido. ■

Combinando los Teoremas 7 y 8 se obtiene el siguiente corolario.

Corolario 9.

Sean $f(x, y)$, $f_1(x)$ y $f_2(y)$ las densidades de probabilidad de (X, Y) , X y Y , respectivamente. Entonces X y Y son independientes si, y solo si,

$$f(x, y) = f_1(x)f_2(y), \quad x, y \in \mathbb{R}.$$

Ejemplo 15.4 La densidad de probabilidad $f(x, y)$ del Ejemplo 15.2 cumple la condición del corolario anterior, por lo tanto las variables aleatorias X y Y de ese ejemplo son independientes. ■

Ejercicio 3.

Pruebe que un vector de variables aleatorias (X, Y) tiene distribución uniforme en el rectángulo $(a, b) \times (c, d)$ si, y solo si, X y Y son independientes y tienen distribución uniforme en los intervalos (a, b) y (c, d) , respectivamente.

Teorema 10.

Sean las variables aleatorias X y Y independientes, con densidades de probabilidad $f_1(x)$ y $f_2(y)$, respectivamente. Entonces $X + Y$ tiene densidad de probabilidad $f(t)$ dada por el producto de convolución de $f_1(x)$ y $f_2(y)$,

$$f(t) = (f_1 * f_2)(t) = \int_{-\infty}^{+\infty} f_2(t - u)f_1(u)du.$$

Demostración:

Sea $F(t)$ la función de distribución de $X + Y$. Utilizando dos veces el teorema

de Fubini se halla que

$$\begin{aligned}
 F(t) &= P(X + Y \leq t) = \iint_{x+y \leq t} f_1(x)f_2(y)dx dy \\
 &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{t-x} f_2(y)dy \right) f_1(x)dx \\
 &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^t f_2(v-x)dv \right) f_1(x)dx \\
 &= \int_{-\infty}^t \left(\int_{-\infty}^{+\infty} f_2(v-x)f_1(x)dx \right) dv.
 \end{aligned}$$

Es decir,

$$F(t) = \int_{-\infty}^t f(v)dv,$$

lo cual demuestra el teorema. ■

Ejemplo 15.5 Sean X y Y variables aleatorias independientes con igual distribución exponencial de parámetro λ . Según el teorema anterior, la densidad de probabilidad $f(t)$ de $X + Y$ se halla,

$$f(t) = \lambda^2 \int_0^t e^{-\lambda(t-u)} e^{-\lambda u} du = \lambda e^{-\lambda t} \lambda t \quad \text{si } t > 0,$$

y $f(t) = 0$ si $t \leq 0$. Luego la distribución de $X + Y$ es Gamma con parámetro de forma $\alpha = 2$ y de escala λ . ■

Similar al caso de una sola variable aleatoria, si (X, Y) es un vector aleatorio con densidad de probabilidad $f(x, y)$ y $g(x, y)$ es una función Boreliana, entonces se tiene que

$$E(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y)dx dy,$$

cuando $g(x, y)$ es no negativa, o la integral anterior converge absolutamente. En particular,

$$EXY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dx dy.$$

El siguiente teorema establece una fórmula, análoga al caso de funciones de una sola variable aleatoria, para obtener la densidad de probabilidad de un vector aleatorio, el cual es función de otro vector aleatorio que tiene densidad de probabilidad.

Teorema 11.

Sea la función $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ tal que $g(x, y)$ es biunívoca, continuamente diferenciable y con Jacobiano diferente de 0. Entonces, dada la densidad de probabilidad $f(x, y)$ del vector (X, Y) , la densidad de probabilidad $h(v, w)$ del vector $(V, W) = g(X, Y)$ se halla por la fórmula

$$h(v, w) = f(g^{-1}(v, w))|J(g^{-1}(v, w))|,$$

donde $|J(g^{-1}(v, w))|$ es el módulo del Jacobiano de la transformación inversa $g^{-1}(v, w)$ de $g(x, y)$.

15.5. Simulación de variables aleatorias con distribución normal

Sean las variables aleatorias X y Y independientes e idénticamente distribuidas $N(0, 1)$. El vector (X, Y) tiene la densidad de probabilidad,

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

Halleemos la densidad de probabilidad $h(v, w)$ del vector aleatorio (V, W) que se obtiene por medio de la transformación de (X, Y) a coordenadas polares. Sean $x = v \cos w$ y $y = v \sin w$, con $v > 0$ y $0 < w \leq 2\pi$. Aplicando el Teorema 11 se tiene,

$$h(v, w) = \frac{1}{2\pi} e^{-v^2/2} v, \quad v > 0, \quad 0 < w \leq 2\pi.$$

No es difícil hallar las densidades marginales de V y W . La variable aleatoria W tiene distribución uniforme en el intervalo $(0, 2\pi)$ y la variable aleatoria V tiene densidad de probabilidad $f_V(v) = v e^{-v^2/2}$ para $v > 0$, y $f_V(v) = 0$ para $v \leq 0$. La distribución de la variable V se denomina *distribución de Rayleigh*. El producto de estas densidades marginales es igual a la

densidad conjunta $h(v, w)$, y por consiguiente las variables aleatorias V y W son independientes.

Puede verificarse que si, inversamente, las variables aleatorias V y W son independientes, V tiene distribución de Rayleigh y W tiene distribución uniforme en el intervalo $(0, 2\pi)$, entonces las variables aleatorias X y Y son independientes y tienen igual distribución $N(0, 1)$. Luego para generar valores de X y Y por medio de la simulación, pueden generarse valores de V y W , y usar las fórmulas,

$$X = V \cos(W), \quad Y = V \sin(W).$$

La variable aleatoria V tiene función de distribución $F_V(v) = 1 - e^{-v^2/2}$ para $v > 0$, cuya función inversa es $F_V^{-1}(y) = \sqrt{-2 \ln(1 - y)}$. Por lo tanto, usando el método estándar de generación de una variable aleatoria, los valores de V se generan por la fórmula $V = \sqrt{-2 \ln(U_1)}$, donde $U_1 \sim U(0, 1)$. Por otro lado, los valores de W se generan usando la fórmula $W = 2\pi U_2$, donde $U_2 \sim U(0, 1)$.

Hemos obtenido las siguientes fórmulas para generar simultáneamente valores de dos variables aleatorias X y Y independientes con distribución $N(0, 1)$:

$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2), \quad Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2),$$

donde U_1 y U_2 son variables aleatorias independientes con idéntica distribución $U(0, 1)$.

Una variable aleatoria X con distribución $N(\mu, \sigma^2)$ puede simularse a partir de la generación de una variable Z con distribución $N(0, 1)$ tomando $X = \sigma Z + \mu$.

Ejercicio 4.

Implemente en lenguaje Python un programa de simulación para generar una muestra de valores de una variable aleatoria con distribución $N(\mu, \sigma^2)$. Calcule la media y la varianza muestral de los valores generados, y compárelos con los valores seleccionados para μ y σ^2 . Construya el histograma de frecuencias correspondiente a la muestra.

15.6. Distribución condicional. Esperanza condicional

Definición 6.

Sea (X, Y) un vector de variables aleatorias discretas con distribución conjunta $P(X = a_i, Y = b_j)$, $i, j \geq 1$. Sea $P(X_i = a_i) > 0$, $i \geq 1$. Se denomina **distribución condicional** de Y dado $X = a_i$ a la colección de probabilidades,

$$P(Y = b_j | X = a_i) = \frac{P(X = a_i, Y = b_j)}{P(X = a_i)}, \quad j \geq 1.$$

De forma análoga se define la distribución condicional de X dado $Y = b_j$. Note que

$$\sum_{j=1}^{\infty} P(Y = b_j | X = a_i) = 1, \quad i \geq 1.$$

Naturalmente que si X y Y son independientes, entonces la distribución condicional de Y dado $X = a_i$ coincide con la distribución marginal de Y .

Ejercicio 5.

Considere las variables aleatorias discretas independientes X y Y . Suponga que $P(Y = y) > 0$. Pruebe que

$$P(X + Y = z | Y = y) = P(X = z - y).$$

Ejercicio 6.

Pruebe que si N_1 y N_2 son variables aleatorias independientes con distribución de Poisson con parámetros λ_1 y λ_2 , respectivamente, entonces $N_1 + N_2$ tiene distribución de Poisson con parámetro $\lambda_1 + \lambda_2$.

Definición 7.

Sea (X, Y) un vector de variables aleatorias discretas con distribución conjunta $P(X = a_i, Y = b_j)$, $i, j \geq 1$. Sean $E|Y| < \infty$ y $P(X = a_i) > 0$, $i \geq 1$. Se denomina **valor esperado condicional de Y dado $X = a_i$** , y se denota por $E(Y|X = a_i)$, a la suma

$$E(Y|X = a_i) = \sum_{j=1}^{\infty} b_j P(Y = b_j|X = a_i).$$

La definición anterior es correcta en el sentido de que el valor esperado condicional existe (y es finito), como muestra la convergencia de la siguiente serie:

$$\begin{aligned} \sum_{j=1}^{\infty} |b_j| P(Y = b_j|X = a_i) &= \sum_{j=1}^{\infty} |b_j| \frac{P(Y = b_j, X = a_i)}{P(X = a_i)} \leq \sum_{j=1}^{\infty} |b_j| \frac{P(Y = b_j)}{P(X = a_i)} \\ &= \frac{E|Y|}{P(X = a_i)} < \infty. \end{aligned}$$

Denotemos $A_i = (X = a_i)$, y sea $P(A_i) > 0$, $i \geq 1$. Puesto que el evento $(Y I_{(X=a_i)} = b_j)$ tiene probabilidad $P(Y = b_j, X = a_i)$, entonces

$$E(Y|X = a_i) = \sum_j b_j \frac{P(Y = b_j, X = a_i)}{P(X = a_i)} = \frac{E(Y I_{(X=a_i)})}{P(X = a_i)} = \frac{E(Y I_{A_i})}{P(A_i)}.$$

Este resultado sugiere la siguiente definición.

Definición 8.

Sean Y una variable aleatoria arbitraria tal que $E|Y| < \infty$, y X una variable aleatoria discreta con valores $\{a_i\}$. Sea $A_i = (X = a_i)$, $P(A_i) > 0$, $i \geq 1$. El **valor esperado condicional de Y dado $X = a_i$** , denotado por $E(Y|X = a_i)$, se define

$$E(Y|X = a_i) = \frac{E(Y I_{A_i})}{P(A_i)}.$$

De la condición $E|Y| < \infty$ en la definición anterior se deduce que $E|YI_{A_i}| < \infty$, lo cual garantiza la existencia de $E(YI_{A_i})$. Note que si X y Y son independientes, entonces $E(Y|X = x) = EY$.

En general, si B es un evento tal que $P(B) > 0$, se define,

$$E(Y|B) = \frac{E(YI_B)}{P(B)}.$$

Luego, en particular,

$$E(I_A|B) = \frac{P(A \cap B)}{P(B)} = P(A|B).$$

Definición 9.

Sean Y una variable aleatoria arbitraria tal que $E|Y| < \infty$, y X una variable aleatoria discreta con valores $a_i, i \geq 1$. Sea $A_i = (X = a_i)$, $P(A_i) > 0, i \geq 1$. Se denomina **valor esperado condicional de Y dado X** a una variable aleatoria, denotada por $E(Y|X)$, definida como

$$E(Y|X) = \sum_{i=1}^{\infty} \frac{E(YI_{A_i})}{P(A_i)} I_{A_i}.$$

Note que la variable aleatoria $E(Y|X)$ toma el valor $E(Y|X = a_i)$ cuando $X = a_i, i \geq 1$.

Teorema 12.

Sean Y una variable aleatoria arbitraria tal que $E|Y| < \infty$, y X una variable aleatoria discreta con valores $a_i, i \geq 1$. Entonces $E|E(Y|X)| < \infty$ y $EY = E[E(Y|X)]$, o sea,

$$EY = \sum_{i=1}^{\infty} E(Y|X = a_i)P(X = a_i). \quad (15.3)$$

Demostración:

Denotando como antes, $A_i = (X = a_i)$, $i \geq 1$. Entonces, como

$$|E(Y|X)| = \sum_{i=1}^{\infty} \frac{|E(YI_{A_i})|}{P(A_i)} I_{A_i},$$

se tiene que

$$\begin{aligned} E|E(Y|X)| &= \sum_{i=1}^{\infty} E\left(\frac{|E(YI_{A_i})|}{P(A_i)} I_{A_i}\right) = \sum_{i=1}^{\infty} |E(YI_{A_i})| \\ &\leq E|Y| \sum_{i=1}^{\infty} I_{A_i} = E|Y| < \infty, \end{aligned}$$

donde la primera igualdad tiene lugar por el corolario del TCM. Si Y es no negativa, esa igualdad sigue siendo válida aun sin los valores absolutos. En el caso en que Y sea arbitraria también es válida por el teorema de Fubini (de la teoría de la medida). ■

Ejemplo 15.6 Supongamos que A es un evento y que $\{B_i\}$ una sucesión de eventos que constituye una partición del espacio muestral tal que $P(B_i) > 0$, $i \geq 1$. Sea X una variable aleatoria definida como $X(\omega) = b_i$ si $\omega \in B_i$, $i \geq 1$. Entonces se obtiene la fórmula de la probabilidad total:

$$\begin{aligned} P(A) &= EI_A = E(E(I_A|X)) = \sum_i E(I_A|X = b_i)P(X = b_i) \\ &= \sum_{i=1}^{\infty} E(I_A|B_i)P(B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i). \end{aligned}$$

■

Ejercicio 7.

Sean N_1 y N_2 variables aleatorias independientes con distribución de Poisson con parámetros λ_1 y λ_2 , respectivamente. Halle

$$E(N_1|N_1 + N_2 = n).$$

Hasta ahora, en esta sección, se asumió que la variable aleatoria Y es integrable y que la variable aleatoria X es discreta. El tratamiento del caso general, en el cual el vector (X, Y) tiene una distribución arbitraria, requiere resultados teóricos que no se consideran en este texto. Por analogía con lo tratado en el caso en que la variable X es discreta, y de manera intuitiva, examinemos el caso en que el vector (X, Y) tiene densidad de probabilidad.

Definición 10.

Supongamos que el vector aleatorio (X, Y) tiene densidad de probabilidad $f(x, y)$, y que la densidad de probabilidad $f_1(x)$ de la variable aleatoria X es diferente de 0 para todo x . Se denomina **densidad de probabilidad condicional** de Y dado $X = x$, y se denota por $f(y|x)$, a la función

$$f(y|x) = \frac{f(x, y)}{f_1(x)}.$$

Es fácil comprobar que

$$\int_{-\infty}^{+\infty} f(y|x) dy = 1.$$

Note que si las variables aleatorias X y Y son independientes, la distribución condicional de Y dado $X = x$ coincide con la distribución marginal de Y .

Definición 11.

Supongamos que el vector aleatorio (X, Y) tiene densidad de probabilidad $f(x, y)$. Sea $E|Y| < \infty$. La **esperanza matemática condicional** de Y dado $X = x$, denotada por $E(Y|X = x)$, se define

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f(y|x) dy.$$

En la siguiente proposición se establece una fórmula análoga a la fórmula (15.3).

Proposición 13.

Supongamos que el vector aleatorio (X, Y) tiene densidad de probabilidad $f(x, y)$. Sea $E|Y| < \infty$. Entonces

$$EY = \int_{-\infty}^{+\infty} E(Y|X = x) f_1(x) dx.$$

Demostración:

Se cumplen las siguientes igualdades:

$$\begin{aligned} EY &= \int_{-\infty}^{+\infty} y f_2(y) dy = \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x, y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{\infty} f(y|x) f_1(x) dx \right) dy \\ &= \int_{-\infty}^{+\infty} f_1(x) \left(\int_{-\infty}^{\infty} y f(y|x) dy \right) dx, \\ &= \int_{-\infty}^{+\infty} E(Y|X = x) f_1(x) dx. \end{aligned}$$

El cambio de orden de las integrales en la penúltima igualdad se justifica por el teorema de Fubini, puesto que $E|Y| < \infty$. ■

Observación 10 Sea $E|Y| < \infty$. En los dos casos tratados, cuando el vector aleatorio (X, Y) tiene densidad de probabilidad, y cuando X es una variable aleatoria discreta, la **varianza condicional de Y dado $X = x$** , denotada por $V(Y|X = x)$, se define como,

$$V(Y|X = x) = E[(Y - E(Y|X = x))^2 | X = x].$$

No es difícil hallar que

$$V(Y|X = x) = E(Y^2 | X = x) - (E(Y|X = x))^2.$$

▲

15.7. Distribución conjunta normal

Definición 12.

Se dice que un vector aleatorio (X, Y) tiene distribución normal, o que las variables aleatorias X y Y tienen **distribución conjunta normal** con parámetros $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, si (X, Y) tiene densidad de probabilidad

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right] \right\}$$

para todos $x, y \in \mathbb{R}$, donde $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ y $-1 < \rho < 1$.

Se denota $(X, Y) \sim N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$. Esta distribución también se denomina **distribución normal bivariada**.

Al hallar las densidades marginales de X y de Y a partir de la densidad conjunta normal se obtiene que $X \sim N(\mu_1, \sigma_1^2)$ y $Y \sim N(\mu_2, \sigma_2^2)$. El cálculo del coeficiente de correlación de X y Y muestra que este es igual al parámetro ρ .

Si en la expresión de la densidad de probabilidad $f(x, y)$ de dos variables aleatorias con distribución conjunta normal se toma $\rho = 0$, inmediatamente se deduce que en este caso las variables son independientes. Luego se obtiene el siguiente resultado simple, pero importante: Dos variables aleatorias con distribución conjunta normal son independientes si, y solo si, son no correlacionadas.

No es difícil hallar la densidad condicional,

$$f(y|x) = \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{1}{2} \left[\frac{y - (\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1))}{\sigma_2\sqrt{1-\rho^2}} \right]^2 \right\}.$$

Note que esta densidad condicional corresponde a una variable aleatoria con distribución normal $N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$. Por consiguiente,

$$E(Y|X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \quad y \quad V(Y|X = x) = \sigma_2^2(1 - \rho^2).$$

La expresión de $E(Y|X = x)$ es la llamada **recta de regresión de Y sobre $X = x$** .

Un método para generar una realización de un vector (X, Y) con distribución $N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$ consiste en lo siguiente: generar un valor de la variable aleatoria X con distribución $N(\mu_1, \sigma_1^2)$; dado que el valor generado de X es x , entonces se genera el valor de Y según la distribución $f(y|x)$, es decir se genera un valor de la distribución $N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$.

Seguidamente se examinará brevemente el caso general de un vector aleatorio $X = (X_1, X_2, \dots, X_n)$, $n \geq 2$, con distribución conjunta normal.

Una forma de caracterizar numéricamente las relaciones entre las variables aleatorias que componen un vector aleatorio es a través de la matriz de covarianzas.

Definición 13.

Se denomina **matriz de covarianzas** de un vector aleatorio $X = (X_1, X_2, \dots, X_n)$, a la matriz cuyo elemento correspondiente a la fila i y columna j es la covarianza $\text{cov}(X_i, X_j)$, $i, j = 1, 2, \dots, n$.

La matriz de covarianzas es una matriz cuadrada, simétrica y definida no negativa. Los elementos de la diagonal principal de la matriz son las varianzas de las variables aleatorias.

Ejercicio 8.

Pruebe que toda matriz de covarianzas es una matriz definida no negativa.

Dada una matriz D , denotemos por D^t su matriz transpuesta. Si D es una matriz cuadrada con determinante diferente de cero, denotaremos su determinante por $\det(D)$. Sean las matrices filas

$$x^t = (x_1 \ x_2 \ \cdots \ x_n) \quad \text{y} \quad \mu^t = (\mu_1 \ \mu_2 \ \cdots \ \mu_n),$$

donde μ_i denota la esperanza matemática de la variable aleatoria X_i , $i = 1, 2, \dots, n$. Denotemos por Σ la matriz de covarianzas del vector aleatorio (X_1, X_2, \dots, X_n) , que supondremos definida positiva.

Definición 14.

Un vector aleatorio $X = (X_1, X_2, \dots, X_n)$ tiene distribución conjunta normal o **distribución normal multivariada**, si tiene la siguiente densidad de probabilidad

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma)}} \exp \left\{ -\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right\}, \quad x \in \mathbb{R}^n.$$

Se denota $X \sim N(\mu, \Sigma)$.

El próximo teorema establece que el vector aleatorio obtenido por medio de una transformación lineal de un vector con distribución normal, también tiene distribución normal.

Por conveniencia consideraremos los vectores de variables aleatorias como matrices columna.

Teorema 14.

Sean $X^t = (X_1 \ X_2 \ \dots \ X_n)$, $X \sim N(\mu, \Sigma)$ y la matriz A de dimensión $m \times n$. Entonces $Y = AX \sim N(A\mu, A\Sigma A^t)$.

Demostración:

La demostración se realizará solo para el caso en que A es una matriz cuadrada con determinante diferente de 0.

Sea n el orden de A . Consideremos primero $X \sim N(0, \Sigma)$. Apliquemos el Teorema 11 a la transformación $y = Ax$. El Jacobiano de la transformación inversa $x = A^{-1}y$ es $1/\det(A)$. Por otro lado,

$$x^t \Sigma^{-1} x = (A^{-1}y)^t \Sigma^{-1} A^{-1}y = y^t (A\Sigma A^t)^{-1} y$$

y, además, $\sqrt{\det(\Sigma)} \det(A) = \sqrt{\det(A\Sigma A^t)}$. Entonces, teniendo en cuenta la Definición 14, se tiene que $Y = AX \sim N(0, A\Sigma A^t)$.

Sea ahora $X \sim N(\mu, \Sigma)$. Por el Teorema 11 es evidente que $X - \mu \sim N(0, \Sigma)$. Notemos que $Y = AX = A(X - \mu) + A\mu$, y que por el resultado ya obtenido, $A(X - \mu) \sim N(0, A\Sigma A^t)$. Utilizando una vez más el Teorema 11, hallamos

finalmente que $Y \sim N(A\mu, A\Sigma A^t)$. ■

Del Teorema 14 se deduce que si $X \sim N(0, I)$, donde I es la matriz identidad, entonces $Y = AX \sim N(0, AA^t)$. Un método para generar una realización de un vector aleatorio Y con distribución $N(0, \Sigma)$, donde Σ es definida positiva, consiste en lo siguiente: usando la *descomposición de Choleski* se obtiene la representación $\Sigma = AA^t$, donde A es una matriz triangular superior y se toma $Y = AX$ con $X \sim N(0, I)$.

Note que cualquier subvector de un vector aleatorio con distribución conjunta normal también tiene distribución conjunta normal. Por ejemplo, si (X, Y, Z) es un vector con distribución conjunta normal, entonces por el Teorema 14 el vector (X, Y) también tiene distribución conjunta normal, pues $(X, Y)^t = A(X, Y, Z)^t$, donde

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

Es fácil deducir el siguiente resultado que generaliza el obtenido anteriormente para dos variables aleatorias con distribución conjunta normal.

Teorema 15.

Si un vector aleatorio $X = (X_1, X_2, \dots, X_n)$ tiene distribución conjunta normal, entonces las variables aleatorias de este vector son independientes si, y solo si, son no correlacionadas.

15.8. Sobre estadística

En esta sección, de forma muy breve, se presentan algunos resultados de gran utilidad en Estadística.

Sean X_1, X_2, \dots, X_n variables aleatorias que constituyen una muestra aleatoria de tamaño n de una variable aleatoria X . Esto quiere decir que las variables aleatorias X_1, X_2, \dots, X_n son independientes e igualmente distribuidas (i.i.d.), con igual distribución que X . Denotemos la varianza muestral por

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

donde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Teorema 16.

Sean X_1, X_2, \dots, X_n variables aleatorias i.i.d. con distribución $N(\mu, \sigma^2)$. Entonces la varianza muestral S^2 y la media muestral \bar{X} son independientes, \bar{X} tiene distribución $N(\mu, \sigma^2/n)$, y $(n-1)S^2/\sigma^2$ tiene distribución $\chi^2(n-1)$.

Demostración:

Como las variables aleatorias X_1, X_2, \dots, X_n son independientes, el vector (X_1, \dots, X_n) tiene distribución conjunta normal. Usando ahora propiedades de la covarianza se halla que

$$\begin{aligned} \text{cov}(X_i - \bar{X}, \bar{X}) &= \text{cov}(X_i, \bar{X}) - \text{cov}(\bar{X}, \bar{X}) = \text{cov}(X_i, \bar{X}) - V(\bar{X}) \\ &= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0, \quad i = 1, 2, \dots, n. \end{aligned}$$

El vector $(X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}, \bar{X})$ se obtiene como una transformación lineal del vector (X_1, \dots, X_n) y, por consiguiente, por el Teorema 14 tiene distribución conjunta normal, luego cada una de sus primeras n componentes y la última son independientes según el Teorema 15, puesto que sus covarianzas son iguales a cero. De aquí se deduce que la media muestral \bar{X} y la varianza muestral S^2 son independientes. Por otro lado, \bar{X} tiene distribución normal, pues se obtiene como una transformación lineal del vector (X_1, \dots, X_n) . Además, $E\bar{X} = \mu$ y $V(\bar{X}) = \sigma^2/n$, de donde $\bar{X} \sim N(\mu, \sigma^2/n)$.

Resta determinar la distribución de S^2 . Probemos la siguiente identidad

$$(n-1)S^2 + n(\bar{X} - \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2. \quad (15.4)$$

Efectivamente,

$$\begin{aligned}
 (n-1)S^2 &= \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2 \\
 &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) \\
 &= \sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) \\
 &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.
 \end{aligned}$$

Dividiendo ambos miembros de la identidad (15.4) por σ^2 se obtiene,

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2.$$

Recordemos ahora que si Z tiene distribución $N(0, 1)$, entonces Z^2 tiene distribución $\chi^2(1)$. Observe que el segundo término de la izquierda tiene distribución $\chi^2(1)$, el término de la derecha tiene distribución $\chi^2(n)$, y los dos sumandos de la izquierda son independientes. Más adelante, cuando se estudie la Parte IV, quedará claro que de esto se deduce que $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución $\chi^2(n-1)$ (vea la Sección 16.3). ■

Ejercicio 9.

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una variable aleatoria X arbitraria con varianza σ^2 . Use la identidad (15.4) para probar que $ES^2 = \sigma^2$.

Los siguientes dos ejercicios se relacionan con la noción de intervalo de confianza. Note que

$$P(|\bar{X} - \mu| < \delta) = P(\bar{X} - \delta < \mu < \bar{X} + \delta).$$

Ejercicio 10.

Considere las variables aleatorias X_1, X_2, \dots, X_n i.i.d con distribución $N(\mu, \sigma^2)$. Halle una fórmula para determinar:

a) Dados σ, n y γ , el valor mínimo de δ tal que

$$P(\bar{X} - \delta < \mu < \bar{X} + \delta) \geq \gamma.$$

b) Dados σ, δ y γ , el valor mínimo de n tal que

$$P(\bar{X} - \delta < \mu < \bar{X} + \delta) \geq \gamma.$$

En la solución del siguiente ejercicio se utilizan resultados del Teorema 16.

Ejercicio 11.

Sean X_1, X_2, \dots, X_n variables aleatorias i.i.d con distribución $N(\mu, \sigma^2)$. Proponga un procedimiento para, dados n y γ , determinar valores de c_1 y c_2 tales que

$$P(c_1(n-1)S^2 < \sigma^2 < c_2(n-1)S^2) \geq \gamma.$$

15.9. Problemas y ejercicios

1. (Ausencia de memoria) Sean X y Y variables aleatorias no negativas e independientes. Suponga que Y tiene densidad de probabilidad, y que X tiene distribución exponencial. Verifique que

$$P(X - Y > x | X > Y) = P(X > x), \quad x > 0.$$

2. Sea $X \sim \text{Gamma}(\lambda, n)$, $n \geq 1$. Pruebe que la función de distribución $F(t)$ de X se expresa

$$F(t) = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \quad t > 0.$$

Sugerencia: Use inducción matemática.

3. Considere n componentes cuyos tiempos de vida son independientes y tienen igual distribución exponencial con parámetro $\lambda = 0,01$ (asuma que la unidad de tiempo es una hora). En el instante inicial comienza a trabajar uno de los componentes y los restantes permanecen de reserva. Cuando se produce el fallo de uno de los componentes, instantáneamente se sustituye por uno de reserva. El fallo del sistema ocurre cuando falla el último de los componentes que cumplía funciones.
 - a) Para $n = 4$ calcule la probabilidad de que el sistema trabaje sin fallo durante 100 h y la probabilidad de que en 50 h se produzcan al menos 2 fallos.
 - b) Calcule el mínimo valor de n tal que la probabilidad de trabajo sin fallo del sistema durante 100 h es mayor que 0,95.
 - c) Utilice la simulación estocástica para calcular aproximadamente la probabilidad de trabajo sin fallo del sistema durante 100 h cuando $n = 4$. Compare los resultados obtenidos con la simulación con el obtenido con la vía analítica en a).
4. Sean X y Y variables aleatorias independientes e igualmente distribuidas con distribución uniforme en el intervalo $[0, 1]$. Halle la distribución de la suma $X + Y$.
5. Sean X y Y variables aleatorias independientes con distribuciones exponenciales con parámetros λ y μ , respectivamente. Halle la densidad de probabilidad de la variable aleatoria $X - Y$. ¿Qué distribución se obtiene cuando $\mu = \lambda$?
6. Considere las variables aleatorias X , Y y Z independientes e igualmente distribuidas con distribución uniforme sobre el intervalo $(0, 1)$. Calcule $P(Z > XY)$.
7. Sean $f_1(x)$, $f_2(y)$ y $f(x, y)$ las densidades de probabilidad de X , Y y (X, Y) , respectivamente. Pruebe que X y Y son independientes si, y solo si, existen funciones $h(x)$ y $g(y)$ tales que $f(x, y) = h(x)g(y)$. (Para variables aleatorias discretas se obtiene un resultado análogo).
8. Considere la siguiente densidad de probabilidad conjunta de (X, Y)

$$f(x, y) = \begin{cases} x + y, & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{en otro caso} \end{cases}$$

- a) Halle la densidad de probabilidad de X .
 - b) Calcule el coeficiente de correlación $\rho(X, Y)$.
 - c) Calcule $P(X + Y < 1)$.
9. Sea (X, Y) un vector de variables aleatorias con distribución uniforme en el círculo con centro en el origen de coordenadas y radio R .
- a) Halle las densidades de probabilidad marginales de X y Y .
 - b) Calcule el coeficiente de correlación. ¿Son X y Y independientes?
 - c) Calcule la probabilidad de que la distancia del origen a (X, Y) sea menor que a .
10. (Intervalo de confianza) Sean X_1, X_2, \dots, X_n variables aleatorias independientes e igualmente distribuidas con distribución exponencial de parámetro λ .
- a) Halle e identifique la densidad de probabilidad de $2\lambda \sum_{i=1}^n X_i$.
 - b) Proponga cómo hallar el menor valor de c tal que, dados γ y n , se cumple:

$$P\left(2\lambda \sum_{i=1}^n X_i < c\right) \geq \gamma.$$

De aquí se obtiene el siguiente intervalo de confianza unilateral para λ :

$$\left(0, \frac{c}{2 \sum X_i}\right).$$

11. Verifique que el cociente de dos variables aleatorias independientes e igualmente distribuidas normales estándar tiene distribución de Cauchy.
12. Sean X y Y variables aleatorias independientes e igualmente distribuidas con distribución $\text{Exp}(\lambda)$. Verifique que $X + Y$ y X/Y son independientes y que la densidad de probabilidad de X/Y es igual a $\frac{1}{(1+x)^2}$, para $x \geq 0$ e igual a 0 para $x < 0$.
13. Considere una sucesión infinita de experimentos independientes de Bernoulli. Sea X la variable aleatoria que representa el número de experimentos hasta el primer éxito. Asuma $EX < \infty$. Condicionando a la ocurrencia de éxito o fracaso en el primer experimento, y usando el valor esperado condicional, calcule EX .

14. *Sean X_1 y X_2 variables aleatorias con distribución binomial de parámetros (n_1, p) y (n_2, p) , respectivamente. Halle $E(X_1|X_1 + X_2 = m)$.
R: $\frac{m}{2}$ cuando $n_1 = n_2$.

15. Sea el vector aleatorio (X, Y) con densidad de probabilidad

$$f(x, y) = 6xy(2 - x - y) \quad \text{si } x, y \in (0, 1),$$

y $f(x, y) = 0$ en otro caso. Halle $E(X|Y = y)$ para $y \in (0, 1)$.

16. Sean X y Y variables aleatorias independientes e igualmente distribuidas con distribución normal. ¿Son $X + Y$ y $X - Y$ dependientes?

17. Considere la siguiente densidad de probabilidad de un vector (X, Y) con distribución normal,

$$f(x, y) = \frac{1}{2\pi\sqrt{3}} \exp \left[-\frac{1}{3}(x^2 - xy + y^2 + x - 2y + 1) \right].$$

Halle: $EX, EY, V(X), V(Y), \rho(X, Y), f(y|x), E(Y|X = x), V(Y|X = x)$.

18. Muestre que con el siguiente procedimiento es posible generar valores de un vector aleatorio (X_1, X_2) con distribución $N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$:

Se consideran Y_1, Y_2 y Y_3 variables aleatorias independientes, cada una con distribución $N(0, 1)$. Sean

$$Z_1 = \sqrt{1 - |c|}Y_1 + \sqrt{|c|}Y_3, \quad Z_2 = \sqrt{1 - |c|}Y_2 \pm \sqrt{|c|}Y_3,$$

donde $c = \sigma_1\sigma_2\rho$, y en la expresión de Z_2 se toma el signo $+$ en el caso $\rho \geq 0$ y el signo $-$ en el caso $\rho \leq 0$. Finalmente se toma,

$$X_1 = \mu_1 + \sigma_1 Z_1, \quad X_2 = \mu_2 + \sigma_2 Z_2.$$

19. Implemente en lenguaje Python los dos métodos descritos en la Sección 15.7, y el método del ejercicio anterior, para generar una muestra de tamaño n de un vector aleatorio (X, Y) con distribución $N(\mu_1, \sigma_1^2; \mu_2, \sigma_2^2; \rho)$, donde $\mu_1 = 1, \sigma_1 = 2, \mu_2 = 4, \sigma_2 = 3$ y $\rho = 0,7$. Considere en el código la visualización gráfica de los puntos generados. Compare los tres métodos.

Parte IV

Función característica. Teoremas límites

FUNCIÓN CARACTERÍSTICA

La función característica tiene gran importancia en el estudio de diversos problemas de la teoría de las probabilidades, en particular de las sumas de variables aleatorias independientes.

En este capítulo las variables aleatorias se considerarán definidas en un mismo espacio de probabilidad.

16.1. Función característica

Definición 1.

Sea $Z = X + iY$, donde X y Y son variables aleatorias con valores reales, e i denota la unidad imaginaria de los números complejos. A Z se le denomina **variable aleatoria compleja**.

Definición 2.

Sean X y Y variables aleatorias integrables. Entonces el valor esperado de la variable aleatoria $Z = X + iY$ se define

$$EZ = EX + iEY.$$

El valor esperado definido sobre el conjunto de las variables aleatorias complejas cumple la propiedad de linealidad. Además, tiene lugar la siguiente proposición.

Proposición 1.

Sea $Z = X + iY$, donde X y Y son variables aleatorias reales e integrables. Entonces

$$|EZ| \leq E|Z| < \infty.$$

Demostración:

Si representamos EZ en la forma polar $EZ = |EZ|e^{i\theta}$, entonces

$$|EZ| = e^{-i\theta} EZ = E\operatorname{Re}(e^{-i\theta} Z) + iE\operatorname{Im}(e^{-i\theta} Z).$$

El miembro izquierdo en las igualdades anteriores es un número real, por consiguiente la parte imaginaria del miembro derecho es igual a 0. Luego, aplicando resultados conocidos sobre el valor esperado de una variable aleatoria real y de operaciones con número complejos, se halla que

$$|EZ| \leq E|\operatorname{Re}(e^{-i\theta} Z)| \leq E|e^{-i\theta} Z| = E|Z|.$$

Por otro lado, $E|Z| < \infty$, puesto que $|Z| = \sqrt{X^2 + Y^2} \leq |X| + |Y|$. ■

Definición 3.

Sea X una variable aleatoria con valores reales. Se denomina **función característica** de X a la función de valores complejos

$$\varphi(t) = Ee^{itX} = E(\cos tX + i \operatorname{sen} tX), \quad t \in \mathbb{R}.$$

Note que la función característica de toda variable aleatoria siempre existe.

A veces se hace referencia a la función característica de una variable aleatoria como la función característica de la distribución de la variable aleatoria. La función característica de una distribución se denomina transformada de Fourier en el Análisis Matemático.

Ejemplo 16.1 a) Si X tiene distribución de Bernoulli con parámetro p se tiene,

$$\varphi(t) = Ee^{itX} = (1-p)e^{it \cdot 0} + pe^{it \cdot 1} = (1-p) + pe^{it}.$$

b) Cuando X tiene distribución exponencial de parámetro λ hallamos,

$$\varphi(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - it} = \frac{1}{1 - \frac{it}{\lambda}}.$$

■

Ejercicio 1.

Halle la función característica de la distribución de Laplace con parámetro λ , $\lambda > 0$, cuya densidad de probabilidad es $f(x) = \frac{\lambda}{2} e^{-\lambda|x|}$, $x \in \mathbb{R}$.

Teorema 2.

Toda función característica $\varphi(t)$ cumple las siguientes propiedades:

1. $|\varphi(t)| \leq 1$, $t \in \mathbb{R}$, y $\varphi(0) = 1$.
2. $\varphi(-t) = \overline{\varphi(t)}$, $t \in \mathbb{R}$.
3. $\varphi(t)$ es uniformemente continua en \mathbb{R} .

Demostración:

La propiedad 1 se deduce inmediatamente usando la Proposición 1. La propiedad 2 es trivial. Demostremos la propiedad 3.

Denotemos por X una variable aleatoria con función de distribución $F(x)$ y función característica $\varphi(t)$. Sean t_1 y t_2 dos números reales arbitrarios y a un número real mayor que 0. Se tiene que

$$\begin{aligned} |\varphi(t_1) - \varphi(t_2)| &\leq E|e^{it_1 X} - e^{it_2 X}| = E|e^{iX(t_1-t_2)} - 1| \\ &\leq E|e^{iX(t_1-t_2)} - 1|I_{(|X| \leq a)} + 2EI_{(|X| > a)} \\ &= I_1 + I_2. \end{aligned}$$

Dado $\epsilon > 0$, tomemos primero a tal que

$$I_2 = 2P(|X| > a) < \frac{\epsilon}{2}.$$

Ahora, fijado a , tomemos $\delta(\epsilon)$ tal que para $|t_1 - t_2| < \delta(\epsilon)$ se cumpla

$$|e^{ix(t_1-t_2)} - 1| < \frac{\epsilon}{2}.$$

Entonces $I_1 < \epsilon/2$, y por lo tanto $|\varphi(t_1) - \varphi(t_2)| < \epsilon$ para cualesquiera t_1 y t_2 tales que $|t_1 - t_2| < \delta(\epsilon)$. ■

Teorema 3.

La función característica de una variable aleatoria determina de manera única su distribución.

Demostración:

En la demostración se considerará el caso de variable aleatoria con densidad de probabilidad continua y derivable por secciones, y el de variable aleatoria discreta.

Cuando la variable aleatoria X tiene densidad de probabilidad $f(x)$, su función característica $\varphi(t)$ se expresa por la fórmula,

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

Si $f(x)$ es continua y derivable por secciones, según la fórmula de inversión de la transformada de Fourier se halla que

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt,$$

lo cual demuestra la propiedad.

Sea ahora X una variable aleatoria discreta con función de probabilidad $f(a_n) = p_n, n \geq 1$, entonces la función característica de X se expresa,

$$\varphi(t) = \sum_{n=1}^{\infty} e^{ita_n} p_n.$$

Multiplicando ambos miembros de la igualdad anterior por e^{-ita_m} se tiene que

$$e^{-ita_m} \varphi(t) = p_m + \sum_{n \neq m} e^{it(a_n - a_m)} p_n,$$

de donde

$$\frac{1}{2T} \int_{-T}^T e^{-ita_m} \varphi(t) dt = p_m + \sum_{n \neq m} p_n \frac{\text{sen } T(a_n - a_m)}{T(a_n - a_m)}.$$

En la integral anterior el cambio de orden con la suma se justifica por la convergencia uniforme de la serie. Tomando límite cuando $T \rightarrow +\infty$, la distribución de la variable aleatoria X se recupera de manera única a partir de su función característica por la fórmula,

$$p_m = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-ita_m} \varphi(t) dt, \quad m \geq 1.$$

■

Ejemplo 16.2 Cuando X es una variable aleatoria de Bernoulli con parámetro p , como $\varphi(t) = (1 - p) + pe^{it}$, entonces la distribución de X se recupera a partir de $\varphi(t)$ de la siguiente manera,

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-it \cdot 1} (1 - p + pe^{it}) dt = p = P(X = 1).$$

$$\lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T e^{-it \cdot 0} (1 - p + pe^{it}) dt = 1 - p = P(X = 0).$$

■

Ejercicio 2.

Se dice que una variable aleatoria X tiene *distribución simétrica* si X y $-X$ tienen igual distribución. Pruebe que una variable aleatoria tiene distribución simétrica si, y solo si, su función característica es una función de valores reales.

El siguiente teorema es de gran utilidad en el estudio de sumas de variables aleatorias independientes.

Teorema 4.

Sean X_1 y X_2 variables aleatorias independientes con funciones características $\varphi_1(t)$ y $\varphi_2(t)$, respectivamente, y $\varphi(t)$ la función característica de $X_1 + X_2$. Entonces

$$\varphi(t) = \varphi_1(t)\varphi_2(t).$$

Demostración:

Puesto que X_1 y X_2 son independientes, se tiene que

$$\varphi(t) = Ee^{it(X_1+X_2)} = E[e^{itX_1}e^{itX_2}] = Ee^{itX_1}Ee^{itX_2} = \varphi_1(t)\varphi_2(t).$$

■

Recordemos el Teorema 10 de la Sección 15.4: si dos variables aleatorias X_1 y X_2 son independientes, y tienen densidad de probabilidad, entonces la densidad de probabilidad de la suma $X_1 + X_2$ es igual al producto de convolución de las densidades de probabilidad de las variables sumandos. Luego del teorema anterior se deduce que la función característica del producto de convolución de dos densidades de probabilidad es igual al producto de sus funciones características.

El teorema 4 se extiende de manera natural para cualquier suma finita de variables aleatorias independientes.

El siguiente ejercicio muestra que la función característica de la suma de variables aleatorias puede ser igual al producto de las funciones características de los sumandos sin ser estos independientes.

Ejercicio 3.

Sean las variables aleatorias $X_1 = X_2$ con distribución de Cauchy estándar. Verifique que

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t).$$

16.2. Función característica de la distribución normal

Proposición 5.

Si $X \sim N(\mu, \sigma^2)$, entonces X tiene función característica

$$\varphi(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

Demostración:

Halleemos primero la función característica $\varphi(t)$ de $Z \sim N(0, 1)$:

$$\varphi(t) = Ee^{itZ} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} dx.$$

Puesto que la integral

$$\int_{-\infty}^{\infty} ix e^{itx - \frac{x^2}{2}} dx$$

converge absoluta y uniformemente para $t \in \mathbb{R}$, y el integrando es una función continua, es posible la diferenciación bajo el símbolo de integral, por lo tanto

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ix e^{itx - \frac{x^2}{2}} dx.$$

Integrando por partes se obtiene,

$$\varphi'(t) = -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} d(e^{-\frac{x^2}{2}}) = -t \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} dx.$$

Es decir, $\varphi(t)$ satisface la ecuación diferencial $\varphi'(t) = -t\varphi(t)$, con $\varphi(0) = 1$. Luego,

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Sea ahora $X \sim N(\mu, \sigma^2)$. Note que X tiene igual distribución que $\sigma Z + \mu$, donde $Z \sim N(0, 1)$. Entonces la función característica de X se halla,

$$\varphi(t) = Ee^{itX} = Ee^{it(\sigma Z + \mu)} = e^{it\mu} Ee^{it\sigma Z} = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

■

Ejemplo 16.3 Sean X_1 y X_2 variables aleatorias independientes con distribución normal, $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Hallemos la distribución de $X = X_1 + X_2$.

La función característica de X se expresa $\varphi(t) = Ee^{itX_1} Ee^{itX_2}$. Sustituyendo las funciones características de X_1 y X_2 , se obtiene

$$\varphi(t) = e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}} = e^{it(\mu_1 + \mu_2) - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}.$$

Luego, $X \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, puesto que la función característica de una variable aleatoria determina de manera única su distribución. ■

16.3. Función característica de la distribución Gamma

Ya se halló que la función característica de una variable aleatoria con distribución exponencial de parámetro λ es igual a $\frac{1}{1 - \frac{it}{\lambda}}$. Consideremos ahora una variable aleatoria $X \sim \text{Gamma}(\lambda, n)$, donde n es un número natural. La función característica de X se expresa

$$\varphi(t) = \int_0^\infty e^{itx} \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x} dx.$$

Integrando por partes, por medio de la inducción matemática se halla que

$$\varphi(t) = \frac{1}{(1 - \frac{it}{\lambda})^n}.$$

Este resultado también se puede obtener usando el Teorema 4, pues X se puede representar como $X = X_1 + X_2 + \dots + X_n$, donde $X_i \sim \text{Exp}(\lambda)$, $i = 1, 2, \dots, n$, son variables aleatorias independientes.

En general, si $X \sim \text{Gamma}(\lambda, \alpha)$, $\alpha > 0$, por un método similar a como se halló la función característica de la distribución normal estándar, se obtiene la función característica de X , la cual tiene la expresión

$$\varphi(t) = \frac{1}{(1 - \frac{it}{\lambda})^\alpha}.$$

Utilizando el Teorema 4 y la unicidad de la función característica, se deduce que si $X_i \sim \text{Gamma}(\lambda, \alpha_i)$, $i = 1, 2$, y si X_1 y X_2 son independientes, entonces $X_1 + X_2 \sim \text{Gamma}(\lambda, \alpha_1 + \alpha_2)$. Note en particular, que si $X_i \sim \chi^2(n_i)$, $i = 1, 2$, entonces $X_1 + X_2 \sim \chi^2(n_1 + n_2)$. Por otro lado, como es conocido, si $X \sim N(0, 1)$, entonces $X^2 \sim \chi^2(1)$. Por consiguiente, tiene lugar la siguiente proposición.

Proposición 6.

Si X_1, X_2, \dots, X_n son variables aleatorias independientes e igualmente distribuidas $N(0, 1)$, entonces la variable aleatoria $\sum_{i=1}^n X_i^2$ tiene distribución $\chi^2(n)$.

Señalemos, finalmente, que cuando X y Y son variables aleatorias independientes tales que X tiene distribución $\chi^2(n)$ y $X + Y$ tiene distribución $\chi^2(n + m)$, entonces Y tiene distribución $\chi^2(m)$, donde $n, m \geq 1$.

16.4. Función característica y momentos

Teorema 7.

Sea $\varphi(t)$ la función característica de la variable aleatoria X . Si X tiene momento de orden n finito, entonces:

1. $\varphi(t)$ tiene n derivadas continuas y

$$\varphi^{(n)}(0) = i^n EX^n.$$

2. Tiene lugar la representación

$$\varphi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} EX^k + \frac{(it)^n}{n!} \varepsilon_n(t),$$

donde $\varepsilon_n(t) \rightarrow 0$ cuando $t \rightarrow 0$.

Demostración:

La demostración se realizará suponiendo que X tiene densidad de probabilidad $f(x)$.

1. Puesto que $E|X| < +\infty$, entonces la integral

$$\int_{-\infty}^{\infty} ixe^{itx} f(x) dx$$

converge absoluta y uniformemente para $t \in \mathbb{R}$, luego es posible la diferenciación de $\varphi(t)$ bajo el símbolo de integral. Por lo tanto,

$$\varphi'(t) = i \int_{-\infty}^{\infty} xe^{itx} f(x) dx, \quad \varphi'(0) = iEX.$$

Además, $\varphi'(t)$ es continua. El resto de esta parte del teorema se demuestra de manera obvia por medio de la inducción matemática.

2. Utilizando la fórmula de Taylor con resto en la forma de Lagrange, se tiene que

$$e^{ix} = \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} + \frac{(ix)^n}{n!} [\cos \theta_1 x + i \sin \theta_2 x],$$

donde $|\theta_1| \leq 1, |\theta_2| \leq 1$. Por lo tanto,

$$e^{itX} = \sum_{k=0}^n \frac{(itX)^k}{k!} + \frac{(itX)^n}{n!} [\cos \theta_1 tX + i \sin \theta_2 tX - 1],$$

donde en la expresión anterior se sumó y restó el término $\frac{(itX)^n}{n!}$. Entonces

$$\varphi(t) = Ee^{itX} = \sum_{k=0}^n \frac{(it)^k}{k!} EX^k + \frac{(it)^n}{n!} \epsilon_n(t),$$

con $\epsilon_n(t) = EX^n(\cos \theta_1 tX + i \sin \theta_2 tX - 1)$. Pero como

$$E|X^n(\cos \theta_1 tX + i \sin \theta_2 tX - 1)| \leq 3E|X|^n < \infty,$$

entonces $\epsilon_n(t)$ converge uniformemente, e intercambiando límite con valor esperado se tiene que $\epsilon_n(t) \rightarrow 0$ cuando $t \rightarrow 0$. ■

El próximo ejemplo muestra la utilidad de la parte 1 del teorema anterior. La parte 2 se utilizará en el siguiente capítulo en la demostración del teorema central del límite.

Ejemplo 16.4 Usando la función característica calculemos la esperanza matemática y la varianza de $X \sim N(\mu, \sigma^2)$.

Todos los momentos de una variable aleatoria con distribución normal son finitos. La función característica de X es

$$\varphi(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

Luego de la parte a) del teorema anterior se obtiene,

$$EX = \frac{\varphi'(0)}{i} = \mu.$$

Además,

$$EX^2 = -\varphi''(0) = \sigma^2 + \mu^2,$$

y por consiguiente, $V(X) = EX^2 - (EX)^2 = \sigma^2$. ■

16.5. Problemas y ejercicios

1. Use la función característica para calcular el valor esperado y la varianza de variables aleatorias con distribuciones Binomial y de Poisson.
2. Sean N_1 y N_2 variables aleatorias independientes con distribución de Poisson. Use la función característica para hallar la distribución de la variable aleatoria $N_1 + N_2$.
3. Suponga que X y Y son variables aleatorias independientes con igual distribución normal estándar. Use la función característica para verificar que $V = (X^2 + Y^2)/2$ tiene distribución exponencial con parámetro igual a 1.

CONVERGENCIAS. TEOREMAS LÍMITES

Los teoremas límites constituyen una parte muy importante de la teoría de las probabilidades.

En las primeras dos secciones de este capítulo las variables aleatorias se considerarán definidas en un mismo espacio de probabilidad.

17.1. Desigualdad de Chebyshev

Recordemos la desigualdad de Markov (Teorema 7 de la Sección 11.2):

Sea X una variable aleatoria no negativa. Entonces para todo $\epsilon > 0$ se tiene

$$P(X \geq \epsilon) \leq \frac{EX}{\epsilon}.$$

Usando la desigualdad de Markov se deduce la desigualdad de Chebyshev.

Teorema 1 (Chebyshev).

Sea X una variable aleatoria arbitraria con valor esperado finito. Entonces

$$P(|X - EX| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2},$$

para todo $\epsilon > 0$.

Demostración:

Observe que

$$P(|X - EX| \geq \epsilon) = P((X - EX)^2 \geq \epsilon^2).$$

Tomemos ahora en la desigualdad de Markov $(X - EX)^2$ en lugar de X , y ϵ^2 en lugar de ϵ . Entonces

$$P((X - EX)^2 \geq \epsilon^2) \leq \frac{E(X - EX)^2}{\epsilon^2} = \frac{V(X)}{\epsilon^2}.$$

■

Ejemplo 17.1 Sea X una variable aleatoria con momento de orden dos finito. Aplicando la desigualdad de Chebyshev se obtiene

$$P(|X - EX| < 3\sigma_X) \geq 1 - 1/9 \approx 0,88.$$

Es decir, la probabilidad de que la diferencia absoluta entre una variable aleatoria y su valor esperado sea menor que tres veces su desviación estándar es no menor que 0,88. Recordemos que si X tiene distribución normal, $P(|X - EX| < 3\sigma_X) \approx 0,997$. ■

17.2. Ley de los grandes números

Definición 1.

Se dice que una sucesión de variables aleatorias $\{X_n\}$ **converge en probabilidad** a la variable aleatoria X si para todo $\epsilon > 0$ se tiene

$$\lim_n P(|X_n - X| > \epsilon) = 0.$$

Se denota $X_n \xrightarrow{p} X$.

Observación 11 La convergencia en probabilidad de la sucesión $\{X_n\}$ a X se puede enunciar de la siguiente manera: para todo $\epsilon > 0$ y todo $\delta > 0$ existe un número natural N tal que si $n \geq N$, entonces $P(|X_n - X| > \epsilon) < \delta$.

Note además que si $\{X_n\}$ converge en probabilidad a X , entonces cualquier subsucesión $\{X_{n'}\}$ de $\{X_n\}$ también converge en probabilidad a X . ▲

Ejercicio 1.

Se dice que una sucesión $\{X_n\}$ **converge en media cuadrática** a la variable aleatoria X si $E(X_n - X)^2 \rightarrow 0$, cuando $n \rightarrow \infty$. Pruebe que la convergencia en media cuadrática de $\{X_n\}$ a X implica la convergencia en probabilidad.

Sea $\{X_n\}$ una sucesión de variables aleatorias. Denotemos $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Los siguientes Teoremas 2 y 3, que establecen resultados en forma de una convergencia en probabilidad, corresponden a la llamada **ley débil de los grandes números**. A la **ley fuerte de los grandes números** corresponden teoremas análogos enunciados de acuerdo con la convergencia casi segura.

Teorema 2 (Chebyshev-Markov).

Sea $\{X_n\}$ una sucesión de variables aleatorias no correlacionadas tales que $EX_i = a$ y $V(X_i) \leq c, i \geq 1$. Entonces

$$\bar{X}_n \xrightarrow{p} a.$$

Demostración:

Puesto que $E\bar{X}_n = a$, aplicando la desigualdad de Chebyshev se obtiene

$$P(|\bar{X}_n - a| > \epsilon) \leq P(|\bar{X}_n - a| \geq \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2}.$$

Como las variables aleatorias son no correlacionadas, entonces

$$V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i),$$

de donde $V(\bar{X}_n) \leq \frac{c}{n}$, y por lo tanto,

$$P(|\bar{X}_n - a| \geq \epsilon) \leq \frac{c}{n\epsilon^2} \rightarrow 0,$$

cuando $n \rightarrow \infty$. ■

Note que el teorema anterior se cumple, en particular, si las variables aleatorias de la sucesión $\{X_n\}$ son independientes e igualmente distribuidas con $EX_1 = a$ y $V(X_1) < \infty$.

El siguiente teorema se deduce de forma inmediata como un corolario del Teorema 2. Lo enunciamos como teorema por constituir el primer resultado teórico de gran importancia en Probabilidades. Fue publicado en el año 1713 en *El Arte de las Conjeturas* de J. Bernoulli.

Teorema 3 (J. Bernoulli).

Sea $\{X_n\}$ una sucesión de variables aleatorias de Bernoulli independientes, con parámetro p . Entonces

$$\bar{X}_n \xrightarrow{p} p.$$

Recordemos que una sucesión de variables aleatorias $\{X_n\}$ converge casi seguramente a una variable aleatoria X , lo cual se denota, $X_n \xrightarrow{c.s.} X$, si $P(X_n \rightarrow X) = 1$.

Teorema 4.

Sean $\{X_n\}$ una sucesión de variables aleatorias y X una variable aleatoria. Se cumple que:

$$\text{Si } X_n \xrightarrow{c.s.} X, \text{ entonces } X_n \xrightarrow{p} X.$$

Demostración:

Tienen lugar las siguientes equivalencias:

$$\begin{aligned} X_n \xrightarrow{c.s.} X &\Leftrightarrow P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} |X_m - X| \leq \frac{1}{k}\right) = 1 \\ &\Leftrightarrow P\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} |X_m - X| \leq \frac{1}{k}\right) = 1, \end{aligned} \quad (17.1)$$

para todo $k \geq 1$.

Pero por la continuidad de la probabilidad, la equivalencia anterior es a su vez equivalente a

$$\lim_n P\left(\bigcap_{m=n}^{\infty} |X_m - X| \leq \frac{1}{k}\right) = 1, \quad (17.2)$$

cualquiera que sea $k \geq 1$. Por otro lado,

$$P\left(\bigcap_{m=n}^{\infty} |X_m - X| \leq \frac{1}{k}\right) \leq P\left(|X_n - X| \leq \frac{1}{k}\right),$$

de donde,

$$\lim_n P\left(|X_n - X| \leq \frac{1}{k}\right) = 1 \quad (17.3)$$

para todo $k \geq 1$. ■

Con palabras, la diferencia entre las convergencias casi segura y en probabilidad de una sucesión de variables aleatorias $\{X_n\}$ a una variable aleatoria X es la siguiente (compare las fórmulas (17.2) y (17.3)): La convergencia casi segura significa que la probabilidad de que *todas* las diferencias absolutas entre los términos de la sucesión X_n y X , a partir de uno de ellos, sean no mayores que un valor dado $\epsilon > 0$, se hace arbitrariamente cercana a 1 tomando el valor de n suficientemente grande. Con la convergencia en probabilidad esto *no necesariamente* ocurre para todas las diferencias, podrían existir diferencias absolutas mayores que ϵ .

Teorema 5.

Sean $\{X_n\}$ una sucesión de variables aleatorias y X una variable aleatoria. Una condición suficiente para que $X_n \xrightarrow{c.s.} X$ es la convergencia de la serie

$$\sum_{m=1}^{\infty} P(|X_m - X| > \epsilon),$$

para todo $\epsilon > 0$.

Demostración:

De acuerdo con (17.1), la convergencia $X_n \xrightarrow{c.s.} X$ equivale a que se cumpla la igualdad,

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} |X_m - X| > \frac{1}{k}\right) = 0$$

para todo $k \geq 1$. Pero por la parte I del Lema de Borel-Cantelli, para que se cumpla esta igualdad es suficiente la condición del enunciado del teorema. ■

Corolario 6.

Si $X_n \xrightarrow{p} X$, entonces existe una subsucesión $\{X_{n'}\}$ de $\{X_n\}$ tal que $X_{n'} \xrightarrow{c.s.} X$.

Demostración:

Tomemos, por ejemplo, X_{n_k} tal que $P(|X_{n_k} - X| > \epsilon) \leq \frac{1}{k^2}$, $k \geq 1$, y apliquemos el Teorema 5. ■

El Ejercicio 7 de la Sección 10.7 establece que si $X_n \xrightarrow{c.s.} X$ y $X_n \xrightarrow{c.s.} Y$, entonces $X = Y$ c.s. En el siguiente ejercicio se establece el resultado análogo con las convergencias en probabilidad.

Ejercicio 2.

Pruebe que si $X_n \xrightarrow{p} X$ y $X_n \xrightarrow{p} Y$, entonces $X = Y$ c.s.

Teorema 7.

Sean $\{X_n\}$ una sucesión de variables aleatorias, X una variable aleatoria y $g(x)$ una función continua. Se cumple que:

1. Si $X_n \xrightarrow{c.s.} X$, entonces $g(X_n) \xrightarrow{c.s.} g(X)$.
2. Si $X_n \xrightarrow{p} X$, entonces $g(X_n) \xrightarrow{p} g(X)$.

Demostración:

1. La demostración se hizo en el Ejemplo 10.11.

2. Supongamos que no tiene lugar la convergencia $g(X_n) \xrightarrow{p} g(X)$. Entonces existen $\epsilon > 0$, $\delta > 0$ y una subsucesión $\{X_{n'}\}$ tales que para todo n' (vea la Observación 11),

$$P(|g(X_{n'}) - g(X)| > \epsilon) > \delta. \quad (17.4)$$

Pero la subsucesión $\{X_{n'}\}$ converge en probabilidad a X y, por consiguiente, según el Corolario 6 existe una subsucesión $\{X_{n''}\}$ de $\{X_{n'}\}$ tal que $X_{n''} \xrightarrow{c.s.} X$, y entonces $g(X_{n''}) \xrightarrow{c.s.} g(X)$, de donde $g(X_{n''}) \xrightarrow{p} g(X)$. Esto contradice la siguiente desigualdad que se deduce de (17.4),

$$P(|g(X_{n''}) - g(X)| > \epsilon) > \delta$$

para todo n'' . ■

El ejemplo que sigue a continuación muestra que una sucesión de variables aleatorias puede converger en probabilidad y no converger casi seguramente.

Ejemplo 17.2 Sea $\{X_n\}$ una sucesión de variables aleatorias independientes tal que

$$P(X_n = 1) = 1/n = 1 - P(X_n = 0), \quad n \geq 1.$$

Puesto que $P(|X_n| > \epsilon) = 0$ si $\epsilon \geq 1$, y $P(|X_n| > \epsilon) = 1/n \rightarrow 0$ si $0 < \epsilon < 1$, entonces $X_n \xrightarrow{p} 0$. Sin embargo, como $\sum_{n=1}^{\infty} P(X_n = 1) = \infty$, por la Parte II del Lema de Borel-Cantelli se tiene que, casi seguramente $X_n = 1$ para infinitos valores de n y, por consiguiente, $\{X_n\}$ no converge casi seguramente a 0. ■

Ejercicio 3.

Sean $\{X_n\}$ y $\{Y_n\}$ sucesiones de variables aleatorias, y X y Y variables aleatorias. Pruebe que:

1. Si $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$, entonces

$$X_n + Y_n \xrightarrow{c.s.} X + Y \quad \text{y} \quad X_n Y_n \xrightarrow{c.s.} XY.$$

2. Si $X_n \xrightarrow{p} X$ y $Y_n \xrightarrow{p} Y$, entonces

$$X_n + Y_n \xrightarrow{p} X + Y \quad \text{y} \quad X_n Y_n \xrightarrow{p} XY.$$

A continuación presentamos un teorema correspondiente a la **ley fuerte de los grandes números**.

Teorema 8 (Kolmogorov).

Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e igualmente distribuidas con valor esperado finito $EX_1 = a$. Entonces

$$\bar{X}_n \xrightarrow{c.s.} a.$$

Demostración:

La demostración se realizará bajo la suposición $EX_1^2 < \infty$. Supongamos primero que las variables aleatorias de la sucesión $\{X_n\}$ son no negativas. De la sucesión $\{\bar{X}_n\}$ tomemos la subsucesión $\{\bar{X}_{k^2}\}$. Se tiene que $V(\bar{X}_{k^2}) = \frac{1}{k^2} V(X_1)$. Sea ahora $\epsilon > 0$. Entonces de la desigualdad de Chebyshev se deduce que

$$\sum_{k=1}^{\infty} P(|\bar{X}_{k^2} - a| > \epsilon) \leq \frac{V(X_1)}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Por lo tanto, por el Teorema 5 se halla que $\bar{X}_{n^2} \xrightarrow{c.s.} a$. Ahora, para n arbitrario, tomemos m , por ejemplo $m = \lfloor \sqrt{n} \rfloor$, tal que $m^2 \leq n \leq (m+1)^2$. Para este m se cumplen las siguientes desigualdades

$$\frac{m^2}{(m+1)^2} \bar{X}_{m^2} \leq \bar{X}_n \leq \bar{X}_{(m+1)^2} \frac{(m+1)^2}{m^2}.$$

De las desigualdades anteriores, como $\bar{X}_{m^2} \xrightarrow{c.s.} a$ y $\bar{X}_{(m+1)^2} \xrightarrow{c.s.} a$ cuando $m \rightarrow \infty$ (y por lo tanto $n \rightarrow \infty$), se concluye que $\bar{X}_n \xrightarrow{c.s.} a$.

Consideremos ahora el caso general. Sea $X_i = X_i^+ - X_i^-$, $i \geq 1$. Denotemos $EX_i^+ = a^+$, $EX_i^- = a^-$, $i \geq 1$, y además,

$$\bar{X}_n^+ = \frac{1}{n} \sum_{i=1}^n X_i^+ \quad \bar{X}_n^- = \frac{1}{n} \sum_{i=1}^n X_i^-.$$

Luego,

$$E\bar{X}_n = E\bar{X}_n^+ - E\bar{X}_n^- = a^+ - a^- = a.$$

Pero si $\bar{X}_n^+ \xrightarrow{c.s.} a^+$ y $\bar{X}_n^- \xrightarrow{c.s.} a^-$, entonces $\bar{X}_n \xrightarrow{c.s.} a^+ - a^- = a$.

■

Recordemos que una variable aleatoria con distribución de Bernoulli de parámetro p tiene todos sus momentos iguales a p . Del teorema anterior se obtiene el siguiente corolario.

Corolario 9.

Sea una sucesión $\{X_n\}$ de variables aleatorias independientes con igual distribución de Bernoulli de parámetro p . Entonces

$$\bar{X}_n \xrightarrow{c.s.} p.$$

En este corolario, \bar{X}_n representa la frecuencia relativa de ocurrencia de éxitos en una sucesión de experimentos independientes, con probabilidad de éxito en cada experimento igual a p . Luego la frecuencia relativa de ocurrencia de éxito converge casi seguramente a su probabilidad p .

Es evidente que del Corolario 9 se deduce el Teorema 3 (de Bernoulli); el primero corresponde a la ley fuerte de los grandes números y el segundo a la ley débil.

Teorema de Bernstein

A continuación se hace una demostración probabilística, debida a S. Bernstein, del conocido teorema de Weierstrass del Análisis Matemático, que establece la existencia de una sucesión de polinomios que converge uniformemente a una función continua en un intervalo cerrado. Es interesante señalar que en el teorema de Bernstein se indica una forma concreta de esta sucesión de polinomios.

Teorema 10 (Bernstein).

Sea $f(x)$ una función real definida y continua en el intervalo $[0, 1]$. Entonces

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \longrightarrow f(x),$$

uniformemente en el intervalo $[0, 1]$ cuando $n \rightarrow \infty$.

Demostración:

Como $f(x)$ es uniformemente continua en $[0, 1]$, cualquiera que sea $\epsilon > 0$ existe $\delta(\epsilon)$ tal si x_1 y x_2 son números reales del intervalo $[0, 1]$ que cumplen $|x_1 - x_2| < \delta$, entonces $|f(x_1) - f(x_2)| < \epsilon$. Por otra parte, $f(x)$ es acotada en $[0, 1]$, luego existe una constante M tal que $|f(x)| \leq M$ para todo $x \in [0, 1]$. Sea ahora $S_n = \sum_{i=1}^n X_i$, donde $\{X_n\}$ es una sucesión de variables de Bernoulli independientes e idénticamente distribuidas, con parámetro x . Entonces, como S_n tiene distribución binomial con parámetros n y x , se tiene que

$$Ef\left(\frac{S_n}{n}\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = B_n(x).$$

Ahora, denotando $\bar{X}_n = S_n/n$, obtenemos

$$\begin{aligned} |B_n(x) - f(x)| &= |Ef(\bar{X}_n) - f(x)| \leq E|f(\bar{X}_n) - f(x)| \\ &= E|f(\bar{X}_n) - f(x)| I_{A_n} + E|f(\bar{X}_n) - f(x)| I_{A_n^c}, \end{aligned}$$

donde $A_n = (|\bar{X}_n - x| \leq \delta)$. De la continuidad uniforme y la acotación de $f(x)$ en $[0, 1]$ se tiene,

$$|B_n(x) - f(x)| \leq \epsilon + 2MEI_{A_n^c} = \epsilon + 2MP(|\bar{X}_n - x| > \delta).$$

Tomando límite cuando $n \rightarrow +\infty$, como \bar{X}_n converge en probabilidad a x , el teorema queda demostrado. ■

Teorema 11 (Convergencia acotada).

Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $|X_n| \leq M$, para todo $n \geq 1$. Entonces $X_n \xrightarrow{p} X$ implica $\lim_n EX_n = EX$.

Demostración:

Probemos primero que $|X| \leq M$ c.s. Para todo $\epsilon > 0$ se cumple:

$$P(|X_n - X| \leq \epsilon) \leq P(|X| - |X_n| \leq \epsilon) = P(|X| \leq |X_n| + \epsilon) \leq P(|X| \leq M + \epsilon).$$

Tomando límite cuando $n \rightarrow \infty$ se tiene que $1 \leq P(|X| \leq M + \epsilon)$. Luego, $P(|X| \leq M + 1/n) = 1$ para todo $n \geq 1$. Por lo tanto,

$$1 = P\left(\bigcap_{n=1}^{\infty} (|X| \leq M + 1/n)\right) = P(|X| \leq M).$$

Sea ahora de nuevo $\epsilon > 0$ arbitrario. Entonces,

$$\begin{aligned} |EX_n - EX| &\leq E|X_n - X| = E|X_n - X|I_{(|X_n - X|) \leq \epsilon} \\ &\quad + E|X_n - X|I_{(|X_n - X|) > \epsilon} \\ &\leq \epsilon P(|X_n - X| \leq \epsilon) + 2MP(|X_n - X| > \epsilon). \end{aligned}$$

Tomando límite cuando $n \rightarrow \infty$ se tiene

$$\limsup_n |EX_n - EX| \leq \epsilon.$$

■

17.3. Convergencia en distribución

Hasta ahora hemos considerado tres tipos de convergencias de una sucesión de variables aleatorias a otra variable aleatoria, definidas todas en el

mismo espacio de probabilidad. Pero puede suceder que ni siquiera sea conocido en qué espacio están definidas las variables aleatorias y, sin embargo, sus distribuciones sean cercanas, recordemos, por ejemplo, la convergencia de la distribución binomial a la distribución de Poisson. En estos casos se trata de la definición de convergencia de una sucesión de funciones de distribución $\{F_n(x)\}$ a una función de distribución $F(x)$.

Se podría sugerir que como medida de proximidad se tome

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|,$$

pero esto no es posible en general, aunque sí en el caso en que $F(x)$ es continua. Sea $X_n = X + 1/n$ y denotemos por $F(x)$ y $F_n(x)$ las funciones de distribución de X y X_n , respectivamente. Entonces $F_n(x) = F(x - 1/n)$, y por lo tanto, $F_n(x) - F(x) = F(x - 1/n) - F(x)$. Si $F(x)$ tiene un punto de salto, $F_n(x)$ no converge a $F(x)$ en ese punto y claro que tampoco es posible utilizar la medida de proximidad sugerida. Sin embargo, es natural considerar que la distribución de X_n converge a la de X en algún sentido.

Definición 2.

Sean $\{X_n\}$ una sucesión de variables aleatorias, X una variable aleatoria y $F(x)$ y $F_n(x)$ las funciones de distribución de X y X_n , $n \geq 1$, respectivamente. Se dice que $\{X_n\}$ **converge en distribución** a X , lo cual se denota $X_n \xrightarrow{d} X$, si

$$\lim_n F_n(x) = F(x),$$

en los puntos de continuidad de $F(x)$.

El próximo teorema establece que si en la definición anterior la función de distribución $F(x)$ es continua, entonces la convergencia es uniforme.

Teorema 12 (Polya).

Sean $\{X_n\}$ una sucesión de variables aleatorias, X una variable aleatoria y $F(x)$ y $F_n(x)$ las funciones de distribución de X y X_n , $n \geq 1$, respectivamente. Si $F(x)$ es continua, entonces $X_n \xrightarrow{d} X$ si, y solo si,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0, \quad \text{cuando } n \rightarrow \infty.$$

Demostración:

La parte *si* es evidente, demostremos la parte *solo si*. Puesto que $F(x)$ es continua, para todo número natural k existen puntos $-\infty = x_0 < x_1 < \dots < x_k = +\infty$ tales que $F(x_i) = i/k$, $i = 0, 1, \dots, k$. Sea ahora $x_{i-1} < x \leq x_i$. Se tiene entonces,

$$F_n(x) - F(x) \leq F_n(x_i) - F(x_{i-1}) = F_n(x_i) - F(x_i) + \frac{1}{k}$$

y

$$F_n(x) - F(x) \geq F_n(x_{i-1}) - F(x_i) = F_n(x_{i-1}) - F(x_{i-1}) - \frac{1}{k}.$$

Luego,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \max_{i=0,1,\dots,k} |F_n(x_i) - F(x_i)| + \frac{1}{k}.$$

Para k dado, como $F_n(x_i) \rightarrow F(x_i)$ para todo $i = 1, 2, \dots, k-1$, cuando $n \rightarrow \infty$, entonces el miembro derecho de la última desigualdad tiende a $\frac{1}{k}$ cuando $n \rightarrow \infty$, pero k se puede tomar todo lo grande que se desee, de donde se obtiene el resultado. ■

El siguiente ejemplo muestra que no toda sucesión de funciones de distribución converge a una función de distribución.

Ejemplo 17.3 Sea $P(X_n = n) = P(X_n = -n) = 1/2$, $n \geq 1$. En este caso la sucesión de funciones de distribución $\{F_n(x)\}$, correspondiente a la sucesión de variables aleatorias $\{X_n\}$, converge a la función $H(x) = 1/2$ para todo x real; pero $H(x)$ no es una función de distribución. ■

Ejercicio 4.

Sea una sucesión de variables aleatorias $\{X_n\}$, y sean las variables aleatorias X y Y con funciones de distribución F y G , respectivamente. Pruebe que si $X_n \xrightarrow{d} X$ y $X_n \xrightarrow{d} Y$, entonces $G \equiv F$.

El próximo teorema muestra que la convergencia en probabilidad implica la convergencia en distribución.

Teorema 13.

Sean $\{X_n\}$ una sucesión de variables aleatorias y X una variable aleatoria. Se cumple que

$$\text{si } X_n \xrightarrow{p} X, \text{ entonces } X_n \xrightarrow{d} X.$$

Demostración:

Denotemos por $F(x)$ y $F_n(x)$, respectivamente, las funciones de distribución de X y X_n , $n \geq 1$. Dado $\epsilon > 0$, denotemos $A_n = (|X_n - X| > \epsilon)$. Entonces para todo x real se tiene que

$$\begin{aligned} P(X_n \leq x) &= P((X_n \leq x) \cap A_n^c) + P((X_n \leq x) \cap A_n) \\ &\leq P(X \leq x + \epsilon) + P(A_n). \end{aligned}$$

Intercambiando X_n y X , y sustituyendo x por $x - \epsilon$ y $x + \epsilon$ por x se obtiene,

$$P(X \leq x - \epsilon) \leq P(X_n \leq x) + P(A_n).$$

De donde,

$$P(X \leq x - \epsilon) - P(A_n) \leq P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(A_n).$$

Como $X_n \xrightarrow{p} X$ significa $\lim_n P(A_n) = 0$, tomando límite cuando $n \rightarrow \infty$ se obtiene,

$$P(X \leq x - \epsilon) \leq \liminf_n P(X_n \leq x) \leq \limsup_n P(X_n \leq x) \leq P(X \leq x + \epsilon).$$

Ahora, si x es un punto de continuidad de $F(x)$, tomado límite cuando $\epsilon \rightarrow 0$, finalmente se halla que $\lim_n F_n(x) = F(x)$. ■

El ejemplo a continuación pone de manifiesto que en general el recíproco del teorema anterior no se cumple.

Ejemplo 17.4 Sea la sucesión $\{X_n\}$, donde $X_n = 1 - X$, $n \geq 1$, y X tiene distribución uniforme en el intervalo $(0, 1)$. Es claro que $X_n \xrightarrow{d} X$. Por otro lado,

$$P(|X_n - X| > \epsilon) = P(|1 - 2X| > \epsilon) = 1 - \epsilon \neq 0, \quad 0 < \epsilon < 1.$$

Por consiguiente, $\{X_n\}$ no converge en probabilidad a X . ■

El recíproco del Teorema 13 se cumple en el caso particular que se considera en el siguiente ejercicio.

Ejercicio 5.

Sea c una constante real. Pruebe que si $X_n \xrightarrow{d} c$, entonces $X_n \xrightarrow{p} c$.

A continuación se presenta un teorema según el cual, si una sucesión de variables aleatorias $\{X_n\}$ converge en distribución a una variable aleatoria X y estas variables aleatorias no están definidas en el mismo espacio de probabilidad, entonces se pueden redefinir en un mismo espacio de probabilidad en el cual $\{X_n\}$ converge a X casi seguramente.

Teorema 14 (Skorohod).

Sean X, X_1, X_2, \dots variables aleatorias con funciones de distribución $F(x), F_1(x), F_2(x), \dots$, respectivamente, tal que $X_n \xrightarrow{d} X$. Entonces existen un espacio de probabilidad, y variables aleatorias Y, Y_1, Y_2, \dots , definidas en este espacio, con funciones de distribución $F(x), F_1(x), F_2(x), \dots$, respectivamente, tal que $Y_n \xrightarrow{c.s.} Y$.

Demostración:

Sea el espacio de probabilidad $([0, 1], B([0, 1]), \lambda)$, y sea la variable aleatoria $U(\omega) = \omega$ definida sobre este. Denotemos por F^{-1} la inversa generalizada de F . Consideremos $Y = F^{-1}(U)$ y $Y_n = F_n^{-1}(U)$. Como $F_n(x) \rightarrow F(x)$ en los puntos de continuidad de $F(x)$, entonces $Y_n(\omega) \rightarrow Y(\omega)$ en los puntos de continuidad de $Y = F^{-1}(\omega)$. Pero los puntos de discontinuidad de $F^{-1}(\omega)$ constituyen a lo más un conjunto infinito numerable. Por consiguiente, $Y_n \xrightarrow{c.s.} Y$. ■

El siguiente teorema es análogo al Teorema 7.

Teorema 15.

Sean $\{X_n\}$ una sucesión de variables aleatorias, X una variable aleatoria y $g(x)$ una función continua. Se cumple que

$$\text{si } X_n \xrightarrow{d} X, \text{ entonces } g(X_n) \xrightarrow{d} g(X).$$

Demostración:

Por el Teorema 14, existen una variable aleatoria Y , y una sucesión de variables aleatorias $\{Y_n\}$, definidas en un mismo espacio de probabilidad, tales que Y tiene igual distribución que X y Y_n igual distribución que X_n , $n \geq 1$, además, $Y_n \xrightarrow{c.s.} Y$. Luego $g(Y_n) \xrightarrow{c.s.} g(Y)$ y, por consiguiente, $g(Y_n) \xrightarrow{d} g(Y)$. Por otro lado, $g(Y)$ tiene igual distribución que $g(X)$ y $g(Y_n)$ igual distribución que $g(X_n)$. Por lo tanto, $g(X_n) \xrightarrow{d} g(X)$. ■

Observe que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$, en general no se cumple que $X_n + Y_n \xrightarrow{d} X + Y$, ni que $X_n Y_n \xrightarrow{d} XY$. Por ejemplo, supongamos que $X_n \xrightarrow{d} X$, donde X tiene distribución $N(0, 1)$; si ahora tomamos $Y_n = -X_n$, $n \geq 1$, entonces también $Y_n \xrightarrow{d} X = Y$, pero $X_n + Y_n \xrightarrow{d} 0$ y $X + Y = 2X$.

Teorema 16 (Slutsky).

Sean $\{X_n\}$ y $\{Y_n\}$ sucesiones de variables aleatorias. Si

$$X_n \xrightarrow{d} X \quad \text{y} \quad Y_n \xrightarrow{p} c,$$

donde c es una constante real, entonces:

$$\text{a) } X_n + Y_n \xrightarrow{d} X + c$$

$$\text{b) } X_n \cdot Y_n \xrightarrow{d} X \cdot c.$$

Demostración:

Se demostrará solo la parte a) del teorema, la parte b) se propone como ejercicio. La parte inicial de esta demostración contiene ideas similares a la demostración del Teorema 13. Denotemos por F_X , F_{X_n} y $F_{X_n+Y_n}$ las funciones de distribución de X , X_n y X_n+Y_n , respectivamente. Dado $\epsilon > 0$, denotemos $A_n = (|Y_n - c| > \epsilon)$. Para todo x real se cumple que:

$$\begin{aligned} F_{X_n+Y_n}(x) &= P(X_n + Y_n \leq x) = P((X_n + Y_n \leq x) \cap A_n^c) \\ &\quad + P((X_n + Y_n \leq x) \cap A_n) \leq P(X_n \leq x - c + \epsilon) + P(A_n) \\ &= F_{X_n}(x - c + \epsilon) + P(A_n). \end{aligned}$$

Supongamos ahora que $x - c$ es un punto de continuidad de F_X . Tomando ϵ suficientemente pequeño, $x - c + \epsilon$ también será un punto de continuidad de F_X , puesto que toda función de distribución tiene a lo más un conjunto numerable de puntos de discontinuidad. Entonces

$$\limsup_n F_{X_n+Y_n}(x) \leq F_X(x - c + \epsilon).$$

De manera análoga se halla que

$$\liminf_n F_{X_n+Y_n}(x) \geq F_X(x - c - \epsilon),$$

si $x - c - \epsilon$ es un punto de continuidad de F_X . Por consiguiente,

$$F_X(x - c - \epsilon) \leq \liminf_n F_{X_n+Y_n}(x) \leq \limsup_n F_{X_n+Y_n}(x) \leq F_X(x - c + \epsilon).$$

Tomando $\epsilon \rightarrow 0$, finalmente concluimos que

$$F_{X_n+Y_n}(x) \rightarrow F_X(x - c) = P(X + c \leq x),$$

cuando $n \rightarrow \infty$. ■

Ejercicio 6.

Pruebe la parte b) del Teorema 16.

No se realizará la demostración del siguiente teorema importante, que permite concluir sobre la convergencia de una sucesión de funciones de distribución a partir de la convergencia de sus correspondientes funciones características.

Teorema 17 (Continuidad).

Sea $\{\varphi_n(t)\}$ la sucesión de funciones características correspondientes a una sucesión de variables aleatorias $\{X_n\}$ y sea $\varphi(t)$ la función característica correspondiente a una variable aleatoria X . Entonces

$$\varphi_n(t) \rightarrow \varphi(t) \quad \text{cuando } n \rightarrow \infty \quad \text{para todo } t,$$

si, y solo si,

$$X_n \xrightarrow{d} X.$$

Aunque el próximo teorema se deduce del Teorema 8, en este caso la demostración se realizará solo bajo la suposición de que el valor esperado de las variables aleatorias de la sucesión es finito.

Teorema 18 (Khinchin).

Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e igualmente distribuidas tal que $EX_1 = a$ es finito. Entonces $\bar{X}_n \xrightarrow{p} a$.

Demostración:

Sea

$$Y_n = \bar{X}_n - a = \frac{1}{n} \sum_{i=1}^n Z_i,$$

donde $Z_i = X_i - a$, $i \geq 1$. Denotemos por $\varphi(t)$ la función característica de Z_i , $i \geq 1$. Entonces la función característica $\varphi_{Y_n}(t)$ de Y_n se expresa,

$$\varphi_{Y_n}(t) = \left[\varphi\left(\frac{t}{n}\right) \right]^n.$$

Por otra parte, como $EZ_1 = 0$, de la segunda parte del Teorema 7 de la Sección 16.4 se tiene que $\varphi(t) = 1 + o(t)$, de donde

$$\varphi_{Y_n}(t) = \left[1 + o\left(\frac{1}{n}\right) \right]^n \rightarrow 1,$$

cuando $n \rightarrow \infty$. Luego por el Teorema 17 se deduce que $Y_n \xrightarrow{d} 0$, y por el Ejercicio 5 finalmente se tiene que $Y_n \xrightarrow{p} 0$. ■

17.4. Teorema Central del Límite

El Teorema Central del Límite, en sus diferentes formas, es uno de los resultados más significativos desde el punto de vista teórico y práctico de la teoría de las probabilidades. A continuación presentamos una de sus formulaciones más simples.

Teorema 19 (Teorema Central del Límite).

Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que $0 < V(X_1) = \sigma^2 < \infty$. Sean $S_n = \sum_{i=1}^n X_i$, $EX_1 = \mu$ y

$$F_n(x) = P\left(\frac{S_n - ES_n}{\sqrt{V(S_n)}} \leq x\right) = P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\right), \quad n \geq 1.$$

Entonces

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \rightarrow 0, \quad \text{cuando } n \rightarrow \infty,$$

donde $\Phi(x)$ es la función de distribución de una variable aleatoria con distribución $N(0, 1)$.

Demostración:

Sea $EX_1 = a$. Se cumple que $ES_n = na$ y $V(S_n) = n\sigma^2$. Denotemos $Z_k = \frac{X_k - a}{\sigma}$. Se tiene que

$$Y_n = \frac{S_n - ES_n}{\sqrt{V(S_n)}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{X_k - a}{\sigma} = \frac{1}{\sqrt{n}} \sum_{k=1}^n Z_k.$$

Observemos ahora que $EZ_k = 0$ y $V(Z_k) = 1$, $k \geq 1$. Luego si denotamos por $\varphi(t)$ la función característica de Z_k , $k \geq 1$, entonces según la segunda parte del Teorema 7 de la Sección 16.4 se deduce que

$$\varphi(t) = 1 - \frac{t^2}{2} + o(t^2),$$

y para la función característica $\varphi_{Y_n}(t)$ de Y_n se cumple,

$$\varphi_{Y_n}(t) = \varphi^n\left(\frac{t}{\sqrt{n}}\right).$$

De esta forma,

$$\lim_n \varphi_{Y_n}(t) = \lim_n \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right]^n = e^{-\frac{t^2}{2}}.$$

Pero la función $e^{-\frac{t^2}{2}}$ es la función característica de una variable aleatoria $N(0, 1)$. Por lo tanto, en virtud de los Teoremas 17 y 12 se obtiene el resultado. ■

Del teorema central del límite se deduce que cuando n es grande, la distribución de $\frac{S_n - ES_n}{\sqrt{V(S_n)}}$ es aproximadamente normal estándar y, por lo tanto, la distribución de la suma S_n es aproximadamente normal **con valor esperado na y varianza $n\sigma^2$** .

El siguiente corolario constituye un caso particular del Teorema Central del límite.

Corolario 20 (Teorema integral de Moivre-Laplace).

Sea $\{X_n\}$ una sucesión de variables aleatorias de Bernoulli independientes, con parámetro p . Denotemos $q = 1 - p$ y $S_n = \sum_{i=1}^n X_i$. Entonces,

$$\lim_n P\left(a < \frac{S_n - np}{\sqrt{npq}} \leq b\right) = \Phi(b) - \Phi(a),$$

donde a y b son números reales, $a < b$, y $\Phi(x)$ es la función de distribución de una variable aleatoria $N(0, 1)$.

Ejemplo 17.5 Consideremos en el esquema de Bernoulli, $n = 1000$ y $p = 1/2$. Sea S_n definida como en el corolario anterior. Como S_n tiene distribución aproximadamente normal con valor esperado np y varianza npq , se tiene que

$$P(450 < S_n \leq 525) \approx \Phi\left(\frac{25}{\sqrt{250}}\right) - \Phi\left(-\frac{50}{\sqrt{250}}\right) \approx 0,95.$$

■

17.5. Problemas y ejercicios

1. Pruebe que si X es una variable aleatoria con valores enteros no negativos, entonces $P(X > 0) \leq EX$.
2. Sea $g(x)$ una función Boreliana positiva no decreciente y X una variable aleatoria arbitraria. Pruebe que para todo $\epsilon > 0$ se cumple la desigualdad,

$$P(X \geq \epsilon) \leq \frac{Eg(X)}{g(\epsilon)}.$$

3. Considere X_1, X_2, \dots variables aleatorias i.i.d. con distribución de Bernoulli con parámetro p . Sea $S_n = \sum_{k=1}^n X_k$. Dado n (dado δ), obtenga δ (obtenga n) tal que

$$P\left(\left|\frac{S_n}{n} - p\right| < \delta\right) \geq 0,95.$$

4. Considere la sucesión de variables aleatorias $\{X_n\}$ tal que

$$P(X_n = 0) = n/(n+1) \quad \text{y} \quad P(X_n = n+1) = 1/(n+1).$$

- a) Pruebe que $X_n \xrightarrow{p} 0$.
- b) Analice la convergencia casi segura de $\{X_n\}$ cuando las variables aleatorias de la sucesión $\{X_n\}$ son independientes.
5. Considere las variables aleatorias X_n y Y_n independientes para cada $n \geq 1$, y las variables aleatorias X y Y independientes. Pruebe que si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$, entonces $X_n + Y_n \xrightarrow{d} X + Y$.
6. Sean $\{X_n\}$ una sucesión de variable aleatorias monótona y X una variable aleatoria. Pruebe que si $X_n \xrightarrow{p} X$, entonces $X_n \xrightarrow{c.s.} X$.
7. Considere el paseo aleatorio simple $\{S_n\}$, donde $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, y las variables aleatorias X_i , $i \geq 1$, son independientes e igualmente distribuidas tal que $P(X_1 = 1) = 1/2 = P(X_1 = -1)$. Sea $Y_n = S_n/\sqrt{n}$. Verifique que

$$\varphi_{Y_n}(t) \rightarrow e^{-t^2/2},$$

cuando $n \rightarrow \infty$.

8. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, tal que $X_n \sim \text{Binomial}(n, p_n)$, $n \geq 1$. Pruebe, usando la función característica, que si $np_n \rightarrow a > 0$ cuando $n \rightarrow \infty$, entonces $X_n \xrightarrow{d} X$, donde $X \sim \text{Poisson}(a)$.
9. Utilizando el Teorema Central del Límite pruebe que:

$$\lim_n e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = 1/2.$$

COMPLEMENTOS

18.1. Teorema de Rényi

El siguiente problema muestra una situación práctica en la cual puede ser útil el teorema de Rényi. Supongamos que determinados artículos deben ser procesados consecutivamente por cierto dispositivo. Denotemos por X_i el tiempo de procesamiento del i -ésimo artículo, $i \geq 1$. Sean X_1, X_2, \dots variables aleatorias i.i.d. En un instante inicial dado se comienza a procesar el primer artículo; la variable aleatoria $S_n = \sum_{i=1}^n X_i$ representa el instante en que finaliza el procesamiento del n -ésimo artículo, $n \geq 1$. Supongamos que durante el procesamiento de cada artículo se detecta si este es defectuoso o no, y que la probabilidad de que un artículo sea defectuoso es p . Denotemos por N el número de artículos procesados hasta que por primera vez se detecta uno defectuoso. El tiempo que transcurre hasta la detección por primera vez de un artículo defectuoso se expresa como la suma aleatoria,

$$S_N = X_1 + X_2 + \cdots + X_N.$$

La variable aleatoria N tiene distribución geométrica con parámetro p y es independiente de las variables aleatorias X_1, X_2, \dots . Denotemos $EX_1 = a$. Note que por la identidad de Wald (Sección 14.1) se tiene,

$$ES_N = EX_1 \cdot EN = \frac{a}{p}.$$

El siguiente teorema establece la distribución asintótica de S_N cuando $p \rightarrow 0$, la cual no depende de la distribución de X_1 .

Teorema 1 (A.Rényi).

Sea $\{X_n\}$ una sucesión de variables aleatorias no negativas, independientes e idénticamente distribuidas, tal que $0 < EX_1 = a < \infty$. Sea N una variable aleatoria con distribución geométrica con parámetro p , e independiente de las variables aleatorias de la sucesión $\{X_n\}$. Denotemos $S = \sum_{i=1}^N X_i$. Entonces

$$\lim_{p \rightarrow 0} P(pS \leq x) = \begin{cases} 1 - e^{-\frac{x}{a}}, & \text{si } x > 0 \\ 0, & \text{si } x \leq 0. \end{cases}$$

Demostración:

Denotemos por $\varphi_{X_1}(t)$, $\varphi_S(t)$ y $\varphi_{pS}(t)$ las funciones características de X_1 , S y pS , respectivamente. Usando la esperanza condicional no es difícil obtener que

$$\begin{aligned} \varphi_S(t) &= \sum_{k=1}^{\infty} E(e^{it(X_1 + \dots + X_N)} | N = k) P(N = k) \\ &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \varphi_{X_1}^k(t) = \frac{p \varphi_{X_1}(t)}{1 - (1-p) \varphi_{X_1}(t)}. \end{aligned}$$

Por lo tanto,

$$\varphi_{pS}(t) = \varphi_S(pt) = \frac{p \varphi_{X_1}(pt)}{1 - (1-p) \varphi_{X_1}(pt)}.$$

De otro lado, por la segunda parte del Teorema 7 de la Sección 16.4 tiene lugar el desarrollo

$$\varphi_{X_1}(pt) = 1 + iapt + o(p),$$

de donde

$$\lim_{p \rightarrow 0} \varphi_{pS}(t) = \lim_{p \rightarrow 0} \frac{p(1 + iapt + o(p))}{1 - (1-p)(1 + iapt + o(p))} = \frac{1}{1 - iat}.$$

Este límite es la función característica de una variable aleatoria con distribución exponencial de parámetro $\lambda = 1/a$. Utilizando ahora el Teorema 17 de

la Sección 17.3 finaliza la demostración. ■

Del teorema de Rényi se deduce la siguiente fórmula aproximada cuando p es muy pequeño,

$$P(S \leq x) \approx 1 - e^{-\frac{p}{a}x}, \quad x > 0.$$

Es decir, del teorema se obtienen condiciones bajo las cuales surge una distribución aproximadamente exponencial, independiente de la distribución de la variable aleatoria X_1 .

Parte V

Nota histórica

En esta nota histórica nos referiremos muy brevemente a trabajos y autores que han ejercido gran influencia en el desarrollo de la teoría de las probabilidades. Por supuesto, muchos otros autores y trabajos, incluso importantes, quedarán sin mencionar.

Un momento significativo en el surgimiento de las bases de la teoría de las probabilidades se encuentra en varias cartas escritas en el año 1654 que forman parte de la correspondencia entre **B. Pascal** (1623-1662, Francia) y **P. Fermat** (1601-1665, Francia). En las cartas, estos científicos investigaron problemas de juegos de azar con dados, que ya habían sido estudiados anteriormente por **L. Pacioli** (1445-1557, Italia) y **M. Tartaglia** (1499-1517, Italia).

Pascal inició la correspondencia con Fermat con el objetivo de confirmar sus propias ideas, y así conjuntamente dieron solución correcta a varios problemas. Pero el concepto de probabilidad estaba todavía ausente, Pascal y Fermat usaban solo la comparación del número de casos favorables a uno u otro suceso de interés. Pascal también promovió el desarrollo de la combinatoria, indicando su significado para la solución de juegos de azar.

La primera publicación dedicada al cálculo de probabilidades, *De los razonamientos en los juegos de azar*, cuyo autor fue **C. Huygens** (1629-1695, Holanda), aparece en el año 1657 como apéndice de un libro, y posteriormente se reeditó en diversas ocasiones. Huygens se interesó en 1655 por la correspondencia entre Pascal y Fermat, y como ellos no habían publicado sus métodos de solución, tuvo él que buscar de forma independiente esos métodos.

La obra de Huygens ejerció gran influencia en **Jacob Bernoulli** (1623-1705, Suiza), a quien se debe la primera gran obra dedicada realmente a la teoría de las probabilidades, *El Arte de las Conjeturas*, publicada en 1713, ocho años después de su muerte, y que constituye el trabajo más original de su autor. Aunque en la fecha de su muerte esta obra estaba inacabada, aún hoy tiene gran significado en la teoría de las probabilidades. En este trabajo Bernoulli prueba la ley de los grandes números en su versión más simple, el llamado teorema de Bernoulli. Con esta obra es que puede considerarse que realmente comienza la historia de la *teoría* de las probabilidades.

A. de Moivre (1667-1754, Francia-Inglaterra) en 1730 dió el segundo gran paso en el desarrollo de la teoría, al demostrar en su *Estudios Analíticos*, para un caso particular, los teoremas local e integral sobre la aproximación de la distribución binomial a la distribución normal. De Moivre también resolvió el famoso problema de la ruina del jugador, que había surgido ya con Huygens.

En el último trabajo de **T. Bayes** (1702-1761, Inglaterra), publicado en 1768, se introducen los conceptos de sucesos excluyentes, sucesos indepen-

dientes y de probabilidad condicional. Además, para dos sucesos se formulan de manera precisa los teoremas sobre la probabilidad de ocurrencia de al menos uno de los sucesos y de la probabilidad de ocurrencia conjunta de ambos sucesos.

G. L. Buffon (1707-1788, Francia) introdujo la noción de probabilidad geométrica y estudió varios problemas relacionados con este concepto, en particular el conocido problema del lanzamiento de una aguja.

P. S. Laplace (1749-1827, Francia) realizó un aporte enorme al desarrollo de la teoría de las probabilidades. En 1812 publicó la primera edición de *Teoría Analítica de Probabilidades*. Esta obra la reeditó en varias oportunidades hasta 1825, ampliándola considerablemente en cada ocasión. El libro expone de manera sistemática todos los conocimientos que existían en esa época acerca de la teoría de las probabilidades y utiliza importantes recursos analíticos. Laplace fue el primero que dio la llamada definición clásica de probabilidad, generalizó los teoremas local e integral de De Moivre e introdujo la amplia utilización de los métodos del análisis matemático.

La teoría de errores de observaciones se construyó en los comienzos del siglo XIX por **A. M. Legendre** (1752-1833, Francia) y **K. F. Gauss** (1777-1855, Alemania) motivada por problemas de astronomía. Como resultado de esto la distribución normal adquirió gran significado, que fue fortalecido además en aquella época por sus aplicaciones a problemas de artillería.

Fue **S. D. Poisson** (1781-1840, Francia) el primero en dar una definición de variable aleatoria, aunque el concepto como tal ya se utilizaba. Poisson halló en 1837 la aproximación de la distribución binomial a la hoy conocida como distribución de Poisson.

En la segunda mitad del siglo XIX se pasó del estudio de los sucesos aleatorios al estudio de las variables aleatorias.

P. L. Chebyshev (1821-1894, Rusia), y sus discípulos **A. A. Markov** (1856-1922, Rusia) y **A. M. Lyapunov** (1857-1918, Rusia) obtuvieron resultados muy importantes en relación con la ley de los grandes números y el teorema central del límite, los cuales suponen la consideración de sumas de variables aleatorias. Chebyshev fue quien primero utilizó amplia y profundamente los conceptos de variable aleatoria y esperanza matemática. Es posible que la primera definición exacta de función de distribución la haya dado Lyapunov en 1900 en uno de sus trabajos.

Durante la segunda mitad del siglo XIX tuvo lugar un gran desarrollo de las ciencias naturales, en particular de la física, y surgieron nuevos proble-

mas, entre ellos problemas relacionados con la física estadística. Se hicieron críticas al concepto de probabilidad. Una de estas críticas la hizo **J. L. Bertrand** (1822-1900, Francia) en un curso de teoría de probabilidades publicado en 1899. Con base en la definición inexacta de probabilidad que existía, Bertrand construyó diversos ejemplos (paradojas) en los cuales parecía que el mismo razonamiento lógico conducía a resultados diferentes en el cálculo de la probabilidad de un suceso.

Hacia finales del siglo XIX era evidente para los grandes científicos la necesidad de fundamentar los conceptos de la teoría de las probabilidades de manera que no pudieran surgir contradicciones motivadas por definiciones imprecisas. El sexto problema de los planteados por **D. Hilbert** (1862-1943, Alemania), en el año 1900, consistía en la construcción axiomática de la teoría de las probabilidades.

Es probable que haya sido **E. F. Borel** (1871-1956, Francia) el primero en indicar la relación de la teoría de las probabilidades con la teoría de conjuntos y la teoría de la medida.

H. L. Lebesgue (1875-1941, Francia) formuló la teoría de la medida en 1901 y al año siguiente hizo la formulación de la integral de Lebesgue, la cual generaliza la integral de Riemann, extendiendo el concepto de área para incluir muchas funciones discontinuas. Sus trabajos conjuntos con Borel iniciaron la teoría moderna de funciones de una variable real.

P. Lévy (1886-1971, Francia) desarrolló la teoría y el uso de las funciones características, estableciendo el llamado teorema de continuidad para funciones características. Obtuvo resultados fundamentales para sumas de variables aleatorias. A partir de 1934 crea e investiga el concepto de martingala, y demuestra el primer teorema de convergencia para martingalas, la llamada ley 0-1. El concepto de variable aleatoria se usaba en los libros de texto en una forma intuitiva, hasta que Lévy en 1925 lo utilizó en un texto suyo en la forma moderna como una función.

Un paso extraordinariamente importante en el desarrollo de la teoría de las probabilidades lo dió **A. N. Kolmogorov** (1903-1987, Rusia) al publicar en 1933 su monografía *Fundamentos de la Teoría de las Probabilidades*, en la cual construye la teoría de las probabilidades en forma rigurosa a partir de axiomas, usando la teoría de la medida. Esto le permitió dar una definición rigurosa de esperanza condicional.

El concepto de proceso estocástico se estableció en el siglo XX. Markov estudió un proceso estocástico constituido por determinado tipo de dependencia en sucesiones de variables aleatorias, las hoy llamadas cadenas de

Markov en tiempo discreto.

N. Wiener (1894-1964, EUA-Suecia) abarcó diferentes áreas de trabajo impulsado por su interés en las aplicaciones. El estudio del movimiento Browniano, también llamado proceso de Wiener, lo introdujo en la teoría de las probabilidades a principios de la década de 1920. Posteriormente estudió procesos estocásticos más generales. Fuera de este campo Wiener es muy conocido por sus trabajos en Cibernética.

En 1931 Kolmogorov publicó el artículo *Métodos analíticos en teoría de las probabilidades*, donde establece los fundamentos de la teoría de los procesos estocásticos de Markov, y en 1934 **A. Y. Khinchin** (1894-1959, Rusia) publicó el trabajo *Teoría de correlación de los procesos estocásticos estacionarios*, en el cual se encuentran las bases de la teoría de los procesos estacionarios. Estos dos trabajos pueden considerarse el comienzo de la construcción de la teoría general de los procesos estocásticos, y constituyeron la fuente de inspiración de una enorme cantidad de trabajos posteriores.

W. Feller (1906-1970, Croacia-EUA) trabajó en la teoría de las probabilidades utilizando la formulación teórica de Kolmogorov. En 1936 publicó *On the theory of stochastic processes*, trabajo muy afín al trabajo de Kolmogorov de 1931. Su trabajo más importante son los dos tomos de *Introduction to Probability Theory and its Applications* (1950-1961), que contienen una exposición muy amplia de la teoría, y numerosos ejemplos y aplicaciones. Feller fue el primer editor ejecutivo de la revista *Mathematical Reviews*.

J. L. Doob (1910-2004, EUA) desarrolló en las décadas de 1940 y 1950 la teoría básica de las martingalas y muchas de sus aplicaciones. En 1953 publicó el libro *Stochastic Processes* que se convirtió en un clásico. Los trabajos de Doob constituyen herramientas muy importantes para el estudio de los procesos estocásticos.

K. Itô (1915-2008, Japón) es el padre del análisis estocástico moderno. En 1942 comenzó a construir el concepto de integral estocástica. Para sus trabajos se basó fundamentalmente en los trabajos de Kolmogorov, Lévy y Doob. La teoría desarrollada por Itô tiene enorme impacto teórico y práctico en aplicaciones a la ingeniería, física y economía. En muchas ciencias no matemáticas se utilizan con frecuencia la famosa fórmula de Itô, la Integral de Itô y las llamadas ecuaciones diferenciales estocásticas según Itô.