

**INTRODUCCIÓN A LA TEORÍA
DE LA PROBABILIDAD, VOL. 2**

Miguel Angel García Alvarez

Prólogo

Hemos oído más de una vez la opinión de que una ciencia debe hallarse edificada sobre conceptos fundamentales, claros y precisamente definidos. En realidad, ninguna ciencia, ni aún la más exacta, comienza por tales definiciones. El verdadero principio de la actividad científica consiste más bien en la descripción de fenómenos, que luego son agrupados, ordenados y relacionados entre sí. Ya en esta descripción se hace inevitable aplicar al material determinadas ideas abstractas extraídas de diversos sectores y, desde luego, no únicamente de la observación del nuevo conjunto de fenómenos descritos... Sólo después de una más profunda investigación del campo de fenómenos de que se trate resulta posible precisar más sus conceptos fundamentales científicos y modificarlos progresivamente, de manera a extender en gran medida su esfera de aplicación, haciéndolos así irrefutables. Éste podrá ser el momento de concretarlos en definiciones. Pero el progreso del conocimiento no tolera tampoco la inalterabilidad de las definiciones. Como nos lo evidencia el ejemplo de la Física, también los “conceptos fundamentales” fijados en definiciones experimentan una perpetua modificación del contenido.

Sigmund Freud

Uno de los conceptos centrales de la Teoría de la Probabilidad es el de variable aleatoria, el cual, en un sentido, generaliza al de evento pues un evento A puede verse como variable aleatoria considerando su función indicadora I_A . Asociadas a un evento, tenemos dos probabilidades, la del evento y la de su complemento. En cambio, asociada a una variable aleatoria X , tenemos una familia de eventos, todos

aquéllos generados por X , a saber, los eventos de la forma $[X \in B]$, en donde B es un subconjunto del conjunto en donde la variable aleatoria toma sus valores.

Dada una variable aleatoria de interés en un determinado problema, el primer objetivo consiste en encontrar la distribución de la variable aleatoria, es decir, el conjunto de probabilidades de los eventos generados por ella. En el caso de las variables aleatorias con valores reales, la distribución de una variable aleatoria X queda determinada por su función de distribución F_X . Dicho de otra manera, la función de distribución contiene toda la información probabilística de la correspondiente variable aleatoria, de manera que esta función adquiere una importancia básica.

La función de distribución F_X de una variable aleatoria es siempre una función monótona no decreciente, de manera que admite una descomposición en una parte de saltos y una parte continua.

Cuando la parte continua de F_X es cero, se dice que la variable aleatoria es discreta y en ese caso F_X queda determinada por la función de densidad f_X de la variable aleatoria, la cual se define, en este caso, mediante la relación $f_X(x) = P[X = x]$. De esta forma, en el caso discreto, el cálculo de cualquier probabilidad se reduce a una suma finita o a una serie, siendo entonces estos conceptos la herramienta matemática que se utiliza para estudiar a la variable aleatoria.

Cuando la parte de saltos de F_X es cero, se dice que la variable aleatoria es continua. En ese caso, es posible que F_X sea no sólo una función continua sino también absolutamente continua, es decir que exista una función f_X tal que $F_X(x) = \int_{-\infty}^x f_X(y)dy$, así que también en este caso la función de distribución queda determinada por una función de densidad. De esta forma, en el caso absolutamente continuo, el cálculo de cualquier probabilidad se reduce a una integral, siendo entonces la Teoría de Integración en una variable la herramienta matemática que se utiliza para estudiar a la variable aleatoria.

No siempre existe una función de densidad asociada a una variable aleatoria, pero la función de distribución siempre existe. En general, una función de distribución representa una medida sobre los conjuntos borelianos, a saber, la medida que comienza por asignar a cada intervalo de la forma $(a, b]$ la medida $F_X(b) - F_X(a)$, y que se extiende después a todos los borelianos. De esta forma, en general, la herramienta que se requiere para el estudio de una variable aleatoria es la Teoría de la Medida en la recta real.

Es muy frecuente que en un problema de probabilidad estemos interesados no en una sola variable aleatoria real, sino en una colección finita de ellas. Esta colección puede verse como un vector aleatorio pues el conjunto de valores que toma una familia de n variables aleatorias es una n -ada de números reales, es decir, un vector en \mathbb{R}^n .

Dado un vector aleatorio de interés en un determinado problema, el primer objetivo consiste en encontrar la distribución de ese vector aleatorio, es decir, el conjunto de

probabilidades de los eventos generados por las variables aleatorias que forman el vector, a saber, los eventos de la forma $[(X_1, \dots, X_n) \in B]$, en donde X_1, \dots, X_n son las variables aleatorias que componen el vector aleatorio y B es un conjunto boreliano de \mathbb{R}^n .

Al igual que en el caso de una sola variable aleatoria, la distribución de un vector aleatorio queda determinada por una función, llamada la función de distribución conjunta de la familia de variables aleatorias que componen al vector. Esta función contiene toda la información probabilística del vector aleatorio.

También en el caso vectorial, la función de distribución conjunta puede ser una función de saltos o una función absolutamente continua, en cuyo caso queda determinada por una función de densidad. La herramienta matemática que se utiliza en estos casos es la Teoría de Series Múltiples, en el caso discreto, y la Teoría de Integrales Múltiples, en el caso absolutamente continuo.

No siempre existe una función de densidad asociada a un vector aleatorio, pero la función de distribución conjunta siempre existe. En general, una función de distribución conjunta de n variables aleatorias representa una medida sobre los conjuntos borelianos de \mathbb{R}^n , de manera que, en general, la herramienta que se requiere para el estudio de un vector aleatorio es la Teoría de la Medida en \mathbb{R}^n .

Un problema de especial interés consiste en determinar el comportamiento en el límite de determinadas funciones definidas mediante familias finitas de variables aleatorias. En este contexto se obtienen los teoremas fundamentales de la Teoría de la Probabilidad, los llamados teoremas límite, entre los que se encuentran la ley débil de los grandes números, la ley fuerte de los grandes números y el teorema del límite central.

Es también frecuente que en un problema de probabilidad estemos interesados no en una colección finita de variables aleatorias reales sino en una infinidad de ellas. En ese caso el tratamiento matemático se complica un poco pues un conjunto de posibles valores de esa familia infinita no puede verse ya como un elemento de un espacio de dimensión finita, a saber, \mathbb{R}^n para alguna $n \in \mathbb{N}$, sino como una función definida sobre el conjunto de índices de la colección infinita de variables aleatorias. Por ejemplo, si tenemos una variable aleatoria X_t para cada número real no negativo t , entonces un conjunto de posibles valores de la familia $(X_t)_{t \geq 0}$ se puede representar mediante una función $f : [0, \infty) \rightarrow \mathbb{R}$, en donde $f(t)$ es el valor que toma la variable aleatoria X_t .

Dada una colección infinita $(X_\gamma)_{\gamma \in \Gamma}$ de variables aleatorias reales, el primer objetivo consiste en encontrar la distribución de esa familia, es decir, el conjunto de probabilidades de los eventos generados por las variables aleatorias que la componen, a saber, los eventos de la forma $[(X_{\gamma_1}, \dots, X_{\gamma_n}) \in B]$, en donde $X_{\gamma_1}, \dots, X_{\gamma_n}$ son elementos de la familia y B es un conjunto boreliano de \mathbb{R}^n de números reales. El problema aquí es que ahora la colección $X_{\gamma_1}, \dots, X_{\gamma_n}$ no es fija.

En este caso, no hay una función de distribución que determine la distribución de la familia infinita de variables aleatorias. Lo que se hace es buscar una medida sobre el conjunto de los posibles valores que toma la familia completa, es decir, sobre el conjunto de funciones $f : \Gamma \rightarrow \mathbb{R}$.

El estudio de la función de distribución de una sola variable aleatoria es un tema que usualmente se desarrolla en un primer curso de probabilidad. Éste se encuentra desarrollado en el primer volumen de este libro.

El estudio de las funciones de distribución conjuntas y, en general, de los vectores aleatorios, incluyendo los teoremas límite, es un tema que usualmente se desarrolla en un segundo curso de probabilidad. Estos temas son el objeto de estudio de este segundo volumen.

El estudio de las familias infinitas de variables aleatorias es un tema que corresponde a la Teoría de Procesos Estocásticos, la cual actualmente es de fundamental importancia en la Teoría de la Probabilidad y sus aplicaciones.

Este segundo volumen integra todo el material que forma parte del programa de un segundo curso de probabilidad que se ofrece en varias universidades.

Al igual que en el primer volumen de este libro, se pretende aquí presentar una introducción a la formulación moderna de la Teoría de la Probabilidad, intentando motivar heurísticamente los conceptos, ubicar el origen de ellos y exponer los resultados con el mayor rigor posible.

Este segundo volumen está dividido en tres grandes partes; en la primera se realiza el estudio de las distribuciones de vectores aleatorios, en la segunda se tratan los teoremas límite y en la tercera se exponen temas sobre la historia de la Teoría de la Probabilidad.

A su vez, la primera parte se divide en cuatro capítulos: en el primero, distribuciones conjuntas, se introduce el concepto de función de distribución conjunta de una familia finita de variables aleatorias; en particular, se tratan los casos discreto y absolutamente continuo, en los cuales existe una función de densidad conjunta. Finalmente se estudia la independencia de variables aleatorias, la cual se caracteriza utilizando la función de distribución conjunta y la densidad conjunta, cuando existe. En el segundo, distribuciones de funciones de vectores aleatorios, se trata el problema consistente en encontrar la distribución de sumas, cocientes, productos y, en general, de cualquier función de una pareja de variables aleatorias; se estudia también el problema consistente en encontrar la distribución conjunta de variables aleatorias definidas como funciones de n variables aleatorias con distribución conjunta conocida; además, se estudian los estadísticos de orden de una familia finita de variables aleatorias absolutamente continuas, independientes e idénticamente distribuidas; finalmente, se tratan problemas relativos al cálculo de esperanzas de funciones de una familia finita de variables aleatorias y se introducen los conceptos de correlación

y covarianza. En el tercer capítulo, distribución normal multivariada, se estudian las transformaciones lineales invertibles de vectores aleatorios formados por variables aleatorias independientes, todas ellas con distribución normal estándar, obteniendo de esta forma lo que se llama la distribución normal multivariada; en particular, se aplican estas ideas para demostrar algunos resultados útiles en la Estadística. En el cuarto capítulo, esperanzas condicionales, se introduce un concepto de especial importancia en la Teoría de la Probabilidad moderna, el de esperanza condicional de una variable aleatoria conocido el valor de otra variable aleatoria; se consideran también las distribuciones condicionales de una variable aleatoria dada otra, tanto en el caso en que las dos son discretas o absolutamente continuas, como en aquél en el cual una es discreta y otra absolutamente continua.

La segunda parte consta de un sólo capítulo: teoremas límite. Se comienza este capítulo estudiando la convergencia de variables aleatorias, introduciendo tres diferentes tipos de convergencia — convergencia en probabilidad, convergencia casi segura y convergencia en distribución — y se estudia la relación que hay entre estos modos de convergencia. Se continúa con el estudio de los teoremas límite, demostrando algunos de los teoremas fundamentales de la Teoría de la Probabilidad: las leyes débil y fuerte de los grandes números y el teorema del límite central; además, se trata el problema de la convergencia de series aleatorias.

Finalmente, la tercera parte se divide en dos capítulos: en el primero, surgimiento del Cálculo de Probabilidades, se analiza principalmente el trabajo realizado, en la Teoría de la Probabilidad, por Blaise Pascal, Pierre de Fermat y Christiaan Huygens, quienes dieron las bases para el desarrollo de un Cálculo de Probabilidades como disciplina matemática independiente. En el segundo capítulo, surgimiento de la Teoría de la Probabilidad moderna, se analiza el proceso que condujo a la formulación axiomática de la Teoría de la Probabilidad, dada por Andrey Nikolaevich Kolmogorov en el año 1933.

Miguel A. García Álvarez
Marzo, 2004

Departamento de Matemáticas
Facultad de Ciencias, UNAM
MÉXICO D.F., 04510
e-mail: magaz@servidor.unam.mx

Notación

$A \cup B$	Unión de los conjuntos A y B
$A \cap B$	Intersección de los conjuntos A y B
$\bigcup_{k=1}^n A_k$	Unión de los conjuntos A_1, \dots, A_n
$\bigcap_{k=1}^n A_k$	Intersección de los conjuntos A_1, \dots, A_n
A^c	Complemento del conjunto A
$A \times B$	Producto cartesiano de los conjuntos A y B
$A \subset B$	El conjunto A está contenido en el conjunto B
$A \supset B$	El conjunto A contiene al conjunto B
\emptyset	Conjunto vacío
\mathbb{N}	Conjunto de los números naturales
\mathbb{Z}	Conjunto de los números enteros
\mathbb{R}	Conjunto de los números reales
$\overline{\mathbb{R}}$	$\mathbb{R} \cup \{-\infty, \infty\}$
$\{n, \dots, m\}$	Conjunto de números enteros entre n y m inclusive
$\{n, n + 1 \dots\}$	Conjunto de números enteros mayores o iguales a n
(a, b)	Intervalo abierto $\{x \in \mathbb{R} \mid a < x < b\}$
$[a, b]$	Intervalo cerrado $\{x \in \mathbb{R} \mid a \leq x \leq b\}$
$(a, b]$	Intervalo semiabierto $\{x \in \mathbb{R} \mid a < x \leq b\}$
$[a, b)$	Intervalo semiabierto $\{x \in \mathbb{R} \mid a \leq x < b\}$
$x \cdot y$	Producto punto de los vectores x y y
$\ x\ $	Norma del vector x
$ x $	Valor absoluto del número real x
$\lceil x \rceil$	Mayor entero menor o igual a x
\bar{z}	Conjugado del número complejo z
$\min(a, b)$	Mínimo entre a y b
$\max(a, b)$	Máximo entre a y b
x^+	$\max(x, 0)$
x^-	$\max(-x, 0)$
$\sum_{k=1}^n x_k$	Suma de los números x_1, \dots, x_n
$\prod_{k=1}^n x_k$	Producto de los números x_1, \dots, x_n
$\ln x$	Logaritmo natural de x
$\binom{n}{k}$	Combinaciones de n elementos tomados de k en k
$g \circ f$	Composición de las funciones f y g
$f : A \mapsto B$	función definida sobre el conjunto A , con valores en el conjunto B
$x \rightsquigarrow \alpha$	x tiende al valor α

Parte 1

VECTORES ALEATORIOS

CAPÍTULO 1

DISTRIBUCIONES CONJUNTAS

Una partícula puede tener una posición o puede tener una velocidad, pero en sentido estricto no puede tener las dos... Cuanto más aclaramos el secreto de la posición, más profundamente se esconde el secreto de la velocidad... Podemos distribuir como queramos la incertidumbre, pero nunca podremos eliminarla.

Werner Heisenberg

Lo que hace que la Naturaleza entrañe contenido probabilístico no es nuestro desconocimiento del mecanismo interno, de las complicaciones internas. La probabilidad parece ser de algún modo intrínseca... Un filósofo dijo una vez: "Para que la ciencia exista, es necesario que las mismas condiciones produzcan siempre los mismos resultados". Pues bien, no los producen.

Richard Phillips Feynman

1.1. Funciones de distribución conjuntas

En todo este volumen se asume que se tiene un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ correspondiente a un determinado experimento aleatorio.

Recordemos además que, dadas las variables aleatorias $X : \Omega \mapsto \mathbb{R}$, $X_1 : \Omega \mapsto \mathbb{R}$, \dots , $X_n : \Omega \mapsto \mathbb{R}$, y los subconjuntos de \mathbb{R} , B, B_1, \dots, B_n , denotamos por $[X \in B]$ al conjunto $\{\omega \in \Omega : X(\omega) \in B\}$ y por $[X_1 \in B_1, \dots, X_n \in B_n]$ a la intersección de los

conjuntos $\{\omega \in \Omega : X_k(\omega) \in B_k\}$, para $k \in \{1, \dots, n\}$. También, si $A \subset \mathbb{R}^n$, denotamos por $[(X_1, \dots, X_n) \in A]$ al conjunto $\{\omega \in \Omega : (X_1(\omega), \dots, X_n(\omega)) \in A\}$.

Toda la información probabilística relativa a una variable aleatoria X está contenida en su función de distribución pues, disponiendo de esta última, se puede obtener la probabilidad de cualquier evento cuya ocurrencia o no ocurrencia dependa del valor que tome X . Dos variables aleatorias pueden ser distintas, vistas como funciones definidas sobre el espacio muestral Ω , pero ser idénticas en cuanto a su distribución y entonces, desde el punto de vista probabilístico, nos dan exactamente la misma información y pueden ser entonces utilizadas indistintamente para el mismo propósito. Por ejemplo, consideremos el experimento aleatorio consistente en elegir al azar un número real en el intervalo $(0, 1)$ y definamos X como el número que se selecciona. El espacio muestral de este experimento es el mismo intervalo $(0, 1)$. Definamos ahora una nueva variable aleatoria Y mediante la fórmula $Y = 1 - X$. Vistas como funciones definidas sobre el espacio muestral, X y Y son diferentes pues por un lado se tiene $X(x) = x$, mientras que por el otro $Y(x) = 1 - x$. Ahora bien, como la elección se realiza al azar, X tiene distribución uniforme en el intervalo $(0, 1)$ y se puede ver inmediatamente que la distribución de Y también es uniforme en el intervalo $(0, 1)$. Por tal motivo, desde el punto de vista probabilístico, X y Y tienen el mismo comportamiento y pueden ser utilizadas indistintamente con el mismo propósito. Por ejemplo, en un problema de simulación, para generar n números que puedan considerarse como provenientes de una variable aleatoria con distribución exponencial de parámetro $\lambda = 1$, se pueden generar n números aleatorios x_1, \dots, x_n cuya distribución sea uniforme en el intervalo $(0, 1)$ y definir, para $k \in \{1, \dots, n\}$, $y_k = -\ln x_k$. Los números y_1, \dots, y_n resuelven entonces el problema planteado. Pero definiendo $z_k = -\ln(1 - x_k)$, los números z_1, \dots, z_n también lo resuelven.

Cuando en un determinado problema son varias las variables aleatorias de interés, la colección de las correspondientes funciones de distribución nos da la información probabilística completa de cada una de las variables aleatorias por separado. Sin embargo, esta colección no nos da la información completa de las variables aleatorias vistas como una familia pues falta la información correspondiente a la posible relación que puede existir entre ellas. Los siguientes 2 ejemplos ilustran este punto.

EJEMPLO 1.1. Consideremos el experimento aleatorio consistente en seleccionar al azar un punto en el interior del cuadrado de vértices $A(0,0)$, $B(1,0)$, $C(1,1)$ y $D(0,1)$. Definamos entonces las variables aleatorias X y Y como la abscisa y ordenada, respectivamente, del punto seleccionado.

Recordemos que en un experimento de este tipo, la probabilidad de que el punto seleccionado esté contenido en un subconjunto A de \mathbb{R}^2 es igual al cociente del área de $A \cap C$ entre el área de C , en donde C representa la región en la cual se selecciona el punto. Con base en esto, las funciones de distribución de X y Y pueden obtenerse fácilmente, llegándose al siguiente resultado:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

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

EJEMPLO 1.2. Consideremos ahora el experimento aleatorio consistente en seleccionar al azar un punto sobre la diagonal de pendiente 1 del cuadrado de vértices $A(0,0)$, $B(1,0)$, $C(1,1)$ y $D(0,1)$. Definamos, como antes, las variables aleatorias X y Y como la abscisa y ordenada, respectivamente, del punto seleccionado.

Recordemos que ahora, la probabilidad de que el punto seleccionado esté contenido en un subconjunto A de \mathbb{R}^2 es igual al cociente de la longitud de $A \cap D$ entre la

longitud de D , en donde D representa la región en la cual se selecciona el punto. Con base en esto, las funciones de distribución de X y Y pueden, nuevamente, obtenerse fácilmente, llegándose al siguiente resultado:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

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

▲

Como puede verse, en ambos problemas se obtienen las mismas funciones de distribución para las variables aleatorias X y Y . Sin embargo, es evidente que la relación entre X y Y es distinta en los dos problemas. En el ejemplo 1.2, conociendo el valor de X se obtiene inmediatamente el de Y pues $Y = X$; en cambio, en el ejemplo 1.1, el conocimiento de X no nos da información sobre el valor de Y pues en cualquier caso éste puede ser cualquier número entre 0 y 1.

En el caso de una familia de n variables aleatorias, el papel central, que juega la función de distribución cuando se trata de una sola variable aleatoria, no lo tiene la colección de las n funciones de distribución correspondientes, sino lo que se llama la función de distribución conjunta, concepto que se define a continuación:

DEFINICIÓN 1.3. [**Función de distribución conjunta**] Sean X_1, \dots, X_n n variables aleatorias. La función $F_{X_1, \dots, X_n} : \mathbb{R}^n \mapsto [0, 1]$, definida por:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

es llamada la función de distribución conjunta de X_1, \dots, X_n .

Para ilustrar esta definición, encontremos la función de distribución conjunta de X y Y en cada uno de los dos ejemplos mencionados previamente.

EJEMPLO 1.4. En el ejemplo 1.1, si $0 < x < 1$ y $0 < y < 1$, entonces $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$ es igual al área de la región sombreada de la figura siguiente:

De manera que se obtiene, $F_{X,Y}(x, y) = xy$.

Considerando los diferentes casos, se obtiene:

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ xy & \text{si } 0 < x < 1, 0 < y < 1 \\ x & \text{si } 0 < x < 1, y \geq 1 \\ y & \text{si } 0 < y < 1, x \geq 1 \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

En el ejemplo 1.2, si $0 < x < 1$ y $0 < y < 1$, entonces $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$ es igual al cociente de la longitud de la región marcada en **negrita** de la figura siguiente, entre la longitud de la diagonal del cuadrado.

De manera que se obtiene, $F_{X,Y}(x, y) = x$.

Considerando los diferentes casos, se obtiene:

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ x & \text{si } 0 < x < 1, y \geq x \\ y & \text{si } 0 < y < 1, y < x \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

Como puede verse, las funciones de distribución conjuntas resultan diferentes. La distinta relación que hay entre X y Y en los dos problemas no queda reflejada en las distribuciones de X y Y por separado, pero si se manifiesta en las distribuciones conjuntas.

▲

La función de distribución conjunta de 2 variables aleatorias nos da la información probabilística de la pareja de variables aleatorias vista como un todo, pero, además, de acuerdo con la siguiente proposición, nos da la información probabilística de cada variable aleatoria por separado, cuyas distribuciones son conocidas como **distribuciones marginales**.

PROPOSICIÓN 1.5. Sean X y Y dos variables aleatorias cualesquiera y sea $F_{X,Y}$ su función de distribución, entonces:

- (i) $\lim_{x \rightsquigarrow \infty} F_{X,Y}(x, y) = F_Y(y)$, para cualquier $y \in \mathbb{R}$.
- (ii) $\lim_{y \rightsquigarrow \infty} F_{X,Y}(x, y) = F_X(x)$, para cualquier $x \in \mathbb{R}$.

Demostración

Sea $y \in \mathbb{R}$ y (x_n) una sucesión monótona creciente de números reales tal que $\lim_{n \rightsquigarrow \infty} x_n = \infty$. Entonces, la sucesión de eventos $[X \leq x_n, Y \leq y]$ es monótona no decreciente y $[Y \leq y] = \bigcup_{n=1}^{\infty} [X \leq x_n, Y \leq y]$, por lo tanto:

$$F_Y(y) = P[Y \leq y] = \lim_{n \rightsquigarrow \infty} P[X \leq x_n, Y \leq y] = \lim_{n \rightsquigarrow \infty} F_{X,Y}(x_n, y)$$

La otra relación se demuestra de manera similar.

■

El resultado se puede extender al caso de n variables aleatorias, al igual que algunas propiedades que tiene la función de distribución de una sola variable aleatoria. Estas propiedades, cuya demostración se deja como ejercicio, se enuncian en la siguiente proposición:

PROPOSICIÓN 1.6. Sean X_1, \dots, X_n n variables aleatorias y sea F_{X_1, \dots, X_n} su función de distribución conjunta, entonces, para cada $(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \in \mathbb{R}^{n-1}$, se tiene:

a) la función $x \mapsto F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$ es no decreciente y continua por la derecha.

$$\begin{aligned} & b) \lim_{x \rightsquigarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) \\ & = F_{X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \end{aligned}$$

$$c) \lim_{x \rightsquigarrow -\infty} F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) = 0$$

Las condiciones arriba mencionadas no son suficientes para que una función F sea una función de distribución. En efecto, consideremos, por ejemplo, la siguiente función:

$$F(x, y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0 \\ x + y & \text{si } x + y < 1, x \geq 0, y \geq 0 \\ 1 & \text{si } x + y \geq 1, x \geq 0, y \geq 0 \end{cases}$$

Esta función tiene las propiedades siguientes:

- (i) Para cada $y \in \mathbb{R}$, la función $x \mapsto F(x, y)$ es no decreciente y continua por la derecha y $\lim_{x \rightsquigarrow -\infty} F(x, y) = 0$.
- (ii) Para cada $x \in \mathbb{R}$, la función $y \mapsto F(x, y)$ es no decreciente y continua por la derecha y $\lim_{y \rightsquigarrow -\infty} F(x, y) = 0$.
- (iii) Las funciones $G : \mathbb{R} \mapsto [0, 1]$ y $H : \mathbb{R} \mapsto [0, 1]$, definidas por $G(y) = \lim_{x \rightsquigarrow \infty} F(x, y)$ y $H(x) = \lim_{y \rightsquigarrow \infty} F(x, y)$, respectivamente, son funciones de distribución en una variable.

Sin embargo, F no es una función de distribución conjunta de alguna pareja de variables aleatorias X, Y . En efecto, si lo fuera, se tendría:

$$P[X \leq x] = \lim_{y \rightsquigarrow \infty} F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$$

$$P[Y \leq y] = \lim_{x \rightsquigarrow \infty} F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y \geq 0 \end{cases}$$

Así que, $P[X = 0] = P[Y = 0] = 1$.

Por lo tanto, se tendría $P[X = 0, Y = 0] = 1$.

Pero, $P[X = 0, Y = 0] \leq F(0, 0) = 0$, lo cual es una contradicción.

En realidad, una función de distribución representa una medida. En el caso de una sola variable se trataría de una medida sobre subconjuntos de números reales. En el caso de la función de distribución de dos variables aleatorias X y Y se trataría de una

medida sobre subconjuntos de \mathbb{R}^2 . En ese caso, dicha medida comenzaría asignando el valor $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$ al rectángulo infinito $(-\infty, x] \times (-\infty, y]$. De manera más general, si $x_1 \leq x_2$ y $y_1 \leq y_2$ y C es el rectángulo $(x_1, x_2] \times (y_1, y_2]$, entonces:

$$P[(X, Y) \in C] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

Así que ese valor sería entonces la medida asignada al rectángulo C .

Obsérvese que, en particular, la cantidad:

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

es no negativa cualquiera que sea la función de distribución $F_{X,Y}$.

Se puede demostrar que basta con que se cumpla esta condición adicional para que una función de dos variables represente la distribución conjunta de dos variables aleatorias. Es decir, se tiene el siguiente resultado:

PROPOSICIÓN 1.7. *Una función $F : \mathbb{R}^2 \mapsto \mathbb{R}$ representa la función de distribución de una pareja de variables aleatorias X, Y si y sólo si se cumplen las siguientes condiciones:*

- (i) *Para cada $y \in \mathbb{R}$, la función $x \mapsto F(x, y)$ es no decreciente y continua por la derecha y $\lim_{x \rightarrow -\infty} F(x, y) = 0$.*
- (ii) *Para cada $x \in \mathbb{R}$, la función $y \mapsto F(x, y)$ es no decreciente y continua por la derecha y $\lim_{y \rightarrow -\infty} F(x, y) = 0$.*
- (iii) *Las funciones $G : \mathbb{R} \mapsto [0, 1]$ y $H : \mathbb{R} \mapsto [0, 1]$, definidas por $G(y) = \lim_{x \rightarrow \infty} F(x, y)$ y $H(x) = \lim_{y \rightarrow \infty} F(x, y)$, respectivamente, son funciones de distribución en una variable.*
- (iv) *Si $x_1 \leq x_2$ y $y_1 \leq y_2$ entonces $F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0$.*

Obsérvese que en el ejemplo considerado arriba se tiene:

$$F(1, 1) - F(0, 1) - F(1, 0) + F(0, 0) = -1$$

Es decir, si F fuera la función de distribución de una pareja de variables aleatoria X, Y y C es el cuadrado $0 < x \leq 1, 0 < y \leq 1$, entonces se tendría $P[(X, Y) \in C] = -1$, lo cual es una contradicción.

De manera general, la función de distribución conjunta, de n variables aleatorias, representa una medida sobre \mathbb{R}^n , y la familia de variables aleatorias X_1, \dots, X_n puede verse como la función de Ω en \mathbb{R}^n que asigna a cada $\omega \in \Omega$ el vector $(X_1(\omega), \dots, X_n(\omega))$;

de esta forma, podemos decir que las variables aleatorias forman un **vector aleatorio** (X_1, \dots, X_n) .

1.2. Funciones de densidad conjuntas

Cuando se estudia la distribución de una variable aleatoria por separado, hay dos casos en los cuales ésta queda determinada por una función de densidad. Nos referimos al caso discreto y al absolutamente continuo. Esta situación puede extenderse al caso de una familia de variables aleatorias, lo cual se desarrolla a continuación. Para claridad en la exposición, primero se trata el caso de una familia formada por dos variables aleatorias y después se enuncian los resultados para el caso general.

DEFINICIÓN 1.8 (Vector aleatorio discreto bidimensional). *Se dice que la pareja de variables aleatorias X, Y forman un vector aleatorio discreto si existe una colección finita o infinita numerable de vectores $(x_1, y_1), (x_2, y_2), \dots$ tales que:*

- (i) $P[X = x_m, Y = y_m] > 0$ para cualquier m
- (ii) $\sum_m P[X = x_m, Y = y_m] = 1$

En este caso, la función $f_{X,Y} : \mathbb{R}^2 \mapsto [0, 1]$ definida por $f_{X,Y}(x, y) = P[X = x, Y = y]$ es llamada la *función de densidad conjunta del vector (X, Y)* .

La propiedad de la aditividad numerable implica la siguiente relación:

$$F_{X,Y}(x, y) = \sum_{\{(u,v) \in \mathbb{R}^2 \mid u \leq x, v \leq y\}} f_{X,Y}(u, v)$$

De manera más general, si $A \subset \mathbb{R}^2$, la propiedad de la aditividad numerable también implica la relación:

$$P[(X, Y) \in A] = \sum_{\{(x,y) \in \mathbb{R}^2 \mid (x,y) \in A\}} f_{X,Y}(x, y)$$

la cual es sumamente útil para calcular probabilidades de eventos cuya ocurrencia depende tanto de los valores de X como de los de Y .

EJEMPLO 1.9. *Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:*

$$f_{X,Y}(x, y) = \begin{cases} cx & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo. Encuentre a) el valor de c , b) $P[X = Y]$, c) $P[X < Y]$ y d) $P[X > Y]$.

Solución

$$a. 1 = \sum_{\{(x,y)|x,y \in \{1,\dots,N\}\}} f_{X,Y}(x,y) = c \sum_{\{(x,y)|x,y \in \{1,\dots,N\}\}} x = cN \sum_{x=1}^N x = c \frac{N^2(N+1)}{2}$$

$$\text{Por lo tanto, } c = \frac{2}{N^2(N+1)}$$

$$b. P[X = Y] = \sum_{\{(x,y)|x,y \in \{1,\dots,N\}, x=y\}} f_{X,Y}(x,y) = c \sum_{\{(x,y)|x,y \in \{1,\dots,N\}, x=y\}} x \\ = c \sum_{x=1}^N x = \frac{2}{N^2(N+1)} \frac{N(N+1)}{2} = \frac{1}{N}$$

$$c. P[X < Y] = \sum_{\{(x,y)|x,y \in \{1,\dots,N\}, x < y\}} f_{X,Y}(x,y) = c \sum_{\{(x,y)|x,y \in \{1,\dots,N\}, x < y\}} x \\ = c \sum_{x=1}^{N-1} \sum_{y=x+1}^N x = c \sum_{x=1}^{N-1} x(N-x) = c \left[N \sum_{x=1}^{N-1} x - \sum_{x=1}^{N-1} x^2 \right] \\ = \frac{2}{N^2(N+1)} \left[N \frac{(N-1)N}{2} - \frac{(N-1)N(2N-1)}{6} \right] = \frac{1}{3} \frac{N-1}{N}$$

$$d. P[X > Y] = 1 - P[X < Y] - P[X = Y] = 1 - \frac{N-1}{3N} - \frac{1}{N} = \frac{2}{3} \frac{N-1}{N}$$

▲

Obsérvese que, en el tipo de problemas que ilustra el último ejemplo, en general, se tendrá que encontrar el valor de una doble sumatoria. Este valor podrá obtenerse ya sea fijando primero x y sumando sobre los valores correspondientes de y , para concluir realizando la sumatoria sobre x , o bien fijando primero y y sumando sobre los valores correspondientes de x , para concluir realizando la sumatoria sobre y . Esto equivale a utilizar alguno de los dos métodos siguientes:

$$P[(X, Y) \in A] = \sum_x P[(X, Y) \in A, X = x] = \sum_x P[(x, Y) \in A, X = x] \\ = \sum_x \sum_{\{y:(x,y) \in A\}} P[X = x, Y = y] \\ P[(X, Y) \in A] = \sum_y P[(X, Y) \in A, Y = y] = \sum_y P[(X, y) \in A, Y = y] \\ = \sum_y \sum_{\{x:(x,y) \in A\}} P[X = x, Y = y]$$

Por ejemplo, la parte c del ejemplo anterior puede también obtenerse de la siguiente manera:

$$P[X < Y] = \sum_{y=1}^N P[X < Y, Y = y] = \sum_{y=1}^N P[X < y, Y = y] \\ = \sum_{y=2}^N \sum_{x=1}^{y-1} P[X = x, Y = y] = c \sum_{y=2}^N \sum_{x=1}^{y-1} x = c \sum_{x=1}^{N-1} \frac{y(y-1)}{2} \\ = \frac{c}{2} \left[\sum_{y=2}^N y^2 - \sum_{y=2}^N y \right] = \frac{c}{2} \left[\sum_{y=1}^N y^2 - \sum_{y=1}^N y \right]$$

$$= \frac{1}{N^2(N+1)} \left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)}{2} \right] = \frac{1}{N} \left[\frac{2N+1}{6} - \frac{1}{2} \right] = \frac{1}{3} \frac{N-1}{N}$$

Evidentemente, el orden que se escoja conduce al mismo resultado, aunque alguno de ellos puede conducir a cálculos más complicados que el otro.

A continuación se trata la extensión de lo hecho arriba para el caso de 2 variables aleatorias al caso de n .

DEFINICIÓN 1.10 (Vector aleatorio discreto n-dimensional). *Se dice que las variables aleatorias X_1, \dots, X_n forman un vector aleatorio discreto si existe una colección finita o infinita numerable de vectores $(x_1^{(1)}, \dots, x_n^{(1)})$, $(x_1^{(2)}, \dots, x_n^{(2)})$, ... tales que:*

- (i) $P \left[X_1 = x_1^{(m)}, \dots, X_n = x_n^{(m)} \right] > 0$, para cualquier m .
- (ii) $\sum_m P \left[X_1 = x_1^{(m)}, \dots, X_n = x_n^{(m)} \right] = 1$

En este caso, la función $f_{X_1, \dots, X_n} : \mathbb{R}^n \mapsto [0, 1]$ definida por:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = P [X_1 = x_1, \dots, X_n = x_n]$$

es llamada la **función de densidad conjunta de $\mathbf{X}_1, \dots, \mathbf{X}_n$** .

La propiedad de la aditividad numerable implica la siguiente relación:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \sum_{\{(y_1, \dots, y_n) | y_1 \leq x_1, \dots, y_n \leq x_n\}} f_{X_1, \dots, X_n}(y_1, \dots, y_n)$$

De manera más general, si $A \subset \mathbb{R}^n$, la propiedad de la aditividad numerable también implica:

$$(1.1) \quad P[(X_1, \dots, X_n) \in A] = \sum_{\{(x_1, \dots, x_n) \in A\}} f_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

EJEMPLO 1.11 (Distribución multinomial). *Consideremos un experimento aleatorio \mathcal{E} que admita únicamente un número finito de posibles resultados, los cuales denotaremos por e_1, \dots, e_r , y sean p_1, \dots, p_r las respectivas probabilidades de que éstos ocurran. Al considerar el experimento aleatorio consistente en la realización de n repeticiones independientes de \mathcal{E} , resulta de interés definir, para cada $k \in \{1, \dots, r\}$, la variable aleatoria N_k como el número de veces en que se obtiene el resultado e_k .*

Para obtener la función de densidad conjunta del vector (N_1, \dots, N_r) , consideremos r números enteros no negativos, n_1, \dots, n_r , tales que $\sum_{k=1}^r n_k = n$.

El espacio muestral, correspondiente al experimento aleatorio consistente en la realización de n repeticiones independientes de \mathcal{E} , consiste de eneadas $(e_{i_1}, \dots, e_{i_n})$, en donde $i_1, \dots, i_n \in \{1, \dots, r\}$, de tal manera que la probabilidad de obtener la eneada $(e_{i_1}, \dots, e_{i_n})$ es igual a $p_{i_1} \cdots p_{i_n}$.

El evento $[N_1 = n_1, \dots, N_r = n_r]$ consta de todas las eneadas tales que, para $k \in \{1, \dots, r\}$, contengan n_k veces el resultado e_k . Cada una de esas eneadas ocurre con probabilidad $p_1^{n_1} \cdots p_r^{n_r}$, así que para calcular la probabilidad del evento:

$$[N_1 = n_1, \dots, N_r = n_r]$$

resta únicamente obtener el número de eneadas que lo componen.

Ahora bien, en n repeticiones de \mathcal{E} , el total de maneras en que se puede obtener, para $k \in \{1, \dots, r\}$, n_k veces el resultado e_k es igual a:

$$\begin{aligned} & \binom{n}{n_1} \binom{n-n_1}{n_2} \cdots \binom{n-n_1-\cdots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots \frac{(n-n_1-\cdots-n_{r-1})!}{n_r!} = \frac{n!}{n_1!n_2!\cdots n_r!} \end{aligned}$$

Se tiene entonces:

$$f_{N_1, \dots, N_r}(n_1, \dots, n_r) = \begin{cases} \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r} & \text{si } \sum_{k=1}^r n_k = n, n_k \in \{0, 1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

DEFINICIÓN 1.12 (Distribución multinomial). Se dice que el vector aleatorio (N_1, \dots, N_r) tiene distribución multinomial con parámetros n, p_1, \dots, p_r si su función de densidad está dada por:

$$f_{N_1, \dots, N_r}(n_1, \dots, n_r) = \begin{cases} \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r} & \text{si } \sum_{k=1}^r n_k = n, n_k \in \{0, 1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

PROPOSICIÓN 1.13. Sea (N_1, \dots, N_r) un vector aleatorio con distribución multinomial de parámetros n, p_1, \dots, p_r . Entonces, dada cualquier subcolección N_{i_1}, \dots, N_{i_s} , tomada de entre las variables aleatorias N_1, \dots, N_r , el vector aleatorio $(N_{i_1}, \dots, N_{i_s}, n - \sum_{j=1}^s N_{i_j})$ tiene distribución multinomial con parámetros $n, p_{i_1}, \dots, p_{i_s}, 1 - \sum_{j=1}^s p_{i_j}$.

Demostración

Obsérvese primero que el vector aleatorio (N_1, \dots, N_r) tiene distribución multinomial con parámetros n, p_1, \dots, p_r si y sólo si $N_r = n - \sum_{k=0}^{r-1} N_k$ y, si n_1, \dots, n_{r-1} son enteros no negativos tales que $\sum_{k=1}^{r-1} n_k \leq n$, entonces:

$$P[N_1 = n_1, \dots, N_{r-1} = n_{r-1}]$$

$$= \frac{n!}{n_1!n_2!\cdots n_{r-1}!(n-n_1-\cdots-n_{r-1})!} p_1^{n_1} \cdots p_{r-1}^{n_{r-1}} (1-p_1-\cdots-p_{r-1})^{n-n_1-\cdots-n_{r-1}}$$

Por otra parte, basta con demostrar el resultado para $s = r - 2$ pues el resultado general, para cualquier $s \in \{1, \dots, r - 2\}$, se obtiene aplicando $r - s - 1$ veces dicho resultado. Además, reordenando la colección N_1, \dots, N_r , se puede asumir que $i_k = k$. De esta manera, para $s = r - 2$ y n_1, \dots, n_{r-2} enteros no negativos tales que $m = \sum_{k=1}^{r-2} n_k \leq n$, se tiene:

$$\begin{aligned} P[N_1 = n_1, \dots, N_{r-2} = n_{r-2}] &= \sum_{n_{r-1}=0}^{n-m} P[N_1 = n_1, \dots, N_{r-1} = n_{r-1}] \\ &= \sum_{n_{r-1}=0}^{n-n_1-\cdots-n_{r-2}} \frac{n!}{n_1!n_2!\cdots n_{r-1}!(n-m-n_{r-1})!} p_1^{n_1} \cdots p_{r-1}^{n_{r-1}} (1-p_1-\cdots-p_{r-1})^{n-m-n_{r-1}} \\ &= \frac{n!}{n_1!\cdots n_{r-2}!} p_1^{n_1} \cdots p_{r-2}^{n_{r-2}} \sum_{n_{r-1}=0}^{n-m} \frac{1}{n_{r-1}!(n-m-n_{r-1})!} p_{r-1}^{n_{r-1}} (1-p_1-\cdots-p_{r-1})^{n-m-n_{r-1}} \\ &= \frac{n!}{n_1!\cdots n_{r-2}!(n-m)!} p_1^{n_1} \cdots p_{r-2}^{n_{r-2}} \sum_{n_{r-1}=0}^{n-m} \frac{(n-m)!}{n_{r-1}!(n-m-n_{r-1})!} p_{r-1}^{n_{r-1}} (1-p_1-\cdots-p_{r-1})^{n-m-n_{r-1}} \\ &= \frac{n!}{n_1!\cdots n_{r-2}!(n-m)!} p_1^{n_1} \cdots p_{r-2}^{n_{r-2}} \sum_{n_{r-1}=0}^{n-m} \binom{n-m}{n_{r-1}} p_{r-1}^{n_{r-1}} (1-p_1-\cdots-p_{r-1})^{n-m-n_{r-1}} \\ &= \frac{n!}{n_1!\cdots n_{r-2}!(n-m)!} p_1^{n_1} \cdots p_{r-2}^{n_{r-2}} (1-p_1-\cdots-p_{r-2})^{n-m} \end{aligned}$$

■

COROLARIO 1.14. *Sea (N_1, \dots, N_r) un vector aleatorio con distribución multinomial de parámetros n, p_1, \dots, p_r , entonces, para $k \in \{1, \dots, r\}$, la variable aleatoria N_k tiene distribución binomial con parámetros n y p_k .*

La definición de función de densidad en el caso n -dimensional absolutamente continuo es similar a la definición en el caso de una sola variable aleatoria.

DEFINICIÓN 1.15 (Vector aleatorio absolutamente continuo). *Se dice que la función de distribución conjunta, F_{X_1, \dots, X_n} , de las variables aleatorias X_1, \dots, X_n es absolutamente continua si existe una función boreliana $f_{X_1, \dots, X_n} : \mathbb{R}^n \mapsto \mathbb{R}$ integrable tal que:*

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_n \cdots dy_1$$

para cualquier vector $(x_1, \dots, x_n) \in \mathbb{R}^n$.

En este caso se dice también que las variables aleatorias X_1, \dots, X_n forman un vector aleatorio absolutamente continuo y la función f_{X_1, \dots, X_n} es llamada una **función de densidad conjunta de $\mathbf{X}_1, \dots, \mathbf{X}_n$** .

Al igual que en el caso de una sola variable aleatoria, cuando existe una función de densidad conjunta de n variables aleatorias, X_1, \dots, X_n , ésta no es única. En efecto,

dada una de ellas, se puede, por ejemplo, modificar su valor en un número finito de puntos y la nueva función que se obtiene sigue siendo una función de densidad conjunta de X_1, \dots, X_n .

Dado un vector $(x_1, \dots, x_n) \in \mathbb{R}^n$, definamos $A = \{(y_1, \dots, y_n) \mid y_1 \leq x_1, \dots, y_n \leq x_n\}$. La propiedad que caracteriza a una función de densidad conjunta de X_1, \dots, X_n se puede escribir entonces de la siguiente manera:

$$(1.2) \quad P[(X_1, \dots, X_n) \in A] = \int_A \cdots \int f_{X_1, \dots, X_n}(y_1, \dots, y_n) dy_1 \cdots dy_n$$

Se puede demostrar que esta misma relación se cumple para cualquier conjunto boreliano $A \subset \mathbb{R}^n$.

EJEMPLO 1.16. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{y} e^{-\frac{x}{y}} e^{-y} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $P[X < 2Y]$ y b) $P[Y < 2X]$.

Solución

$$\begin{aligned} a. \quad P[X < 2Y] &= \int_0^\infty \int_0^{2y} \frac{1}{y} e^{-\frac{x}{y}} e^{-y} dx dy = \int_0^\infty e^{-y} \int_0^{2y} \frac{1}{y} e^{-\frac{x}{y}} dx dy \\ &= \int_0^\infty e^{-y} (1 - e^{-2}) dy = 1 - e^{-2} \end{aligned}$$

$$\begin{aligned} b. \quad P[Y < 2X] &= \int_0^\infty \int_{\frac{y}{2}}^\infty \frac{1}{y} e^{-\frac{x}{y}} e^{-y} dx dy = \int_0^\infty e^{-y} \int_{\frac{y}{2}}^\infty \frac{1}{y} e^{-\frac{x}{y}} dx dy \\ &= \int_0^\infty e^{-y} e^{-\frac{1}{2}} dy = e^{-\frac{1}{2}} \end{aligned}$$

▲

La función de distribución conjunta de las variables aleatorias X y Y del ejemplo 1.1 es absolutamente continua. En efecto, la función:

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad conjunta de X y Y .

En cambio, la función de distribución conjunta de las variables aleatorias X y Y del ejemplo 1.2 no es absolutamente continua. En efecto, si existiera una función de densidad conjunta, $f_{X,Y}$, de X y Y , se tendría $1 = P[(X, Y) \in D] = \iint_D f_{X,Y}(x, y) dy dx = 0$, ya que el área de D es cero. Obsérvese que en este caso la función de distribución conjunta $F_{X,Y}$ es una función continua y que, vistas por separado, tanto X como Y son absolutamente continuas.

En general, se puede afirmar que si la pareja (X, Y) toma únicamente valores dentro de un subconjunto A de \mathbb{R}^2 de área cero, entonces la función de distribución $F_{X,Y}$ no es absolutamente continua pues en caso de serlo se tendría:

$$1 = P[(X, Y) \in A] = \iint_A f_{X,Y}(x, y) dy dx = 0.$$

EJEMPLO 1.17. *Un experimento aleatorio consiste en seleccionar al azar un punto sobre la base AB de un triángulo equilátero ABC cada uno de cuyos lados mide 2 unidades. Sean X y Y las distancias del punto seleccionado a los vértices C y A , respectivamente. ¿Existe una función de densidad conjunta de X y Y ?*

C

A

B

Solución

No existe una función de densidad conjunta pues la pareja (X, Y) únicamente toma valores sobre la parte de la hipérbola $x^2 - (y - 1)^2 = 3$ que se encuentre dentro del rectángulo $\sqrt{3} \leq x \leq 2, 0 \leq y \leq 2$, la cual tiene un área igual a cero.

▲

En el caso de un experimento aleatorio consistente en la elección al azar de un punto dentro de un subconjunto R de \mathbb{R}^2 , definamos las variables aleatorias X y Y como la abscisa y ordenada, respectivamente, del punto seleccionado. Si el área de la región R está bien definida y no es cero, entonces la función $f : \mathbb{R}^2 \mapsto \mathbb{R}$ definida por:

$$f(x, y) = \begin{cases} \frac{1}{\text{Área de } R} & \text{si } (x, y) \in R \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad conjunta de X y Y . En efecto, si A es un subconjunto de \mathbb{R}^2 para el cual la integral $\iint_A f(x, y) dx dy$ está bien definida, entonces:

$$P[(X, Y) \in A] = \frac{\text{Área de } A \cap R}{\text{Área de } R} = \iint_A f(x, y) dx dy$$

De una manera más general, en el caso de un experimento aleatorio consistente en la elección al azar de un punto dentro de un subconjunto R de \mathbb{R}^n , definamos las variables aleatorias X_1, \dots, X_n como las coordenadas del punto seleccionado. Si la integral $\int \dots \int_R f(x_1, \dots, x_n) dx_1 \dots dx_n$ está bien definida y es positiva, entonces la función $f : \mathbb{R}^n \mapsto \mathbb{R}$ definida por:

$$f(x_1, \dots, x_n) = \begin{cases} \frac{1}{\int \dots \int_R f(x_1, \dots, x_n) dx_1 \dots dx_n} & \text{si } (x_1, \dots, x_n) \in R \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad conjunta de X_1, \dots, X_n . En efecto, si $A \subset \mathbb{R}^n$ y la integral $\int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$ existe, entonces:

$$P[(X_1, \dots, X_n) \in A] = \frac{\int \dots \int_{A \cap R} f(x_1, \dots, x_n) dx_1 \dots dx_n}{\int \dots \int_R f(x_1, \dots, x_n) dx_1 \dots dx_n}$$

Por otra parte, al igual que en el caso de una sola variable aleatoria, cuando la función de distribución, F_{X_1, \dots, X_n} , de una familia de variables aleatorias, X_1, \dots, X_n , es derivable, entonces es absolutamente continua y la función de densidad conjunta se obtiene mediante la fórmula siguiente:

$$f_{X_1, \dots, X_n} = \frac{\partial^n F_{X_1, \dots, X_n}}{\partial x_1 \dots \partial x_n}$$

De manera más específica, la continuidad absoluta de un vector aleatorio n -dimensional puede establecerse aplicando el siguiente resultado:

PROPOSICIÓN 1.18. *Sea (X_1, \dots, X_n) un vector aleatorio con función de distribución conjunta F_{X_1, \dots, X_n} y $S \subset \mathbb{R}^n$ un conjunto abierto tal que $P[(X_1, \dots, X_n) \in S] = 1$. Supongamos que:*

- (i) F_{X_1, \dots, X_n} es continua sobre \mathbb{R}^n .
- (ii) $\frac{\partial^n F_{X_1, \dots, X_n}}{\partial x_1 \dots \partial x_n}$ existe y es continua sobre S .

Entonces, el vector aleatorio (X_1, \dots, X_n) es absolutamente continuo y la función:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \begin{cases} \frac{\partial^n F_{X_1, \dots, X_n}}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n) & \text{si } (x_1, \dots, x_n) \in S \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad conjunta, f_{X_1, \dots, X_n} .

EJEMPLO 1.19. Consideremos la función de distribución del ejemplo 1.1, la cual está dada por:

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ xy & \text{si } 0 < x < 1, 0 < y < 1 \\ x & \text{si } 0 < x < 1, y \geq 1 \\ y & \text{si } 0 < y < 1, x \geq 1 \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

Se puede ver inmediatamente que $F_{X,Y}$ es continua sobre \mathbb{R}^2 y, si $S = (0, 1) \times (0, 1)$, entonces $P[(X, Y) \in S] = 1$ y $\frac{\partial^2 F_{X,Y}}{\partial x \partial y}$ existe y es continua sobre S . Por lo tanto, el vector aleatorio (X, Y) es absolutamente continuo y una función de densidad conjunta, $f_{X,Y}$, de (X, Y) está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) & \text{si } (x, y) \in S \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} 1 & \text{si } (x, y) \in S \\ 0 & \text{en otro caso} \end{cases}$$

EJEMPLO 1.20. Consideremos la función de distribución del ejemplo 1.2, la cual está dada por:

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ x & \text{si } 0 < x < 1, y \geq x \\ y & \text{si } 0 < y < 1, y < x \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

Se puede ver inmediatamente que $F_{X,Y}$ es continua sobre \mathbb{R}^2 y si $S = (0, 1) \times (0, 1)$, entonces $P[(X, Y) \in S] = 1$. Pero $\frac{\partial^2 F_{X,Y}}{\partial x \partial y}$ no existe sobre S pues, para $0 < x < 1$, vista como función de y , $\frac{\partial F_{X,Y}}{\partial x}$ es discontinua en $y = x$.

Por lo tanto, no se puede concluir, basándonos en la proposición 1.18, que la pareja X, Y forme un vector aleatorio absolutamente continuo.

EJEMPLO 1.21. Sea (X, Y) un vector aleatorio con función de distribución conjunta $F_{X,Y}$ dada por:

$$F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ \frac{1}{2}(x^3y + xy^3) & \text{si } 0 < x < 1, 0 < y < 1 \\ \frac{1}{2}(x^3 + x) & \text{si } 0 < x < 1, y \geq 1 \\ \frac{1}{2}(y + y^3) & \text{si } 0 < y < 1, x \geq 1 \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

Encuentre $P[2X < 3Y]$.

Solución

Se puede ver inmediatamente que $F_{X,Y}$ es continua sobre \mathbb{R}^2 y si $S = (0, 1) \times (0, 1)$, entonces $P[(X, Y) \in S] = 1$ y $\frac{\partial^2 F_{X,Y}}{\partial x \partial y}$ existe y es continua sobre S . Por lo tanto, el vector aleatorio (X, Y) es absolutamente continuo y una función de densidad conjunta, $f_{X,Y}$, de (X, Y) está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x, y) & \text{si } (x, y) \in S \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{3}{2}(x^2 + y^2) & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

De manera que:

$$P[2X < 3Y] = \frac{3}{2} \int_0^1 \int_{\frac{2x}{3}}^1 (x^2 + y^2) dy dx = \frac{77}{108}$$

1.3. Funciones de densidad marginales

Sea (X_1, \dots, X_n) un vector aleatorio discreto con función de densidad conjunta f_{X_1, \dots, X_n} . Por la propiedad de la aditividad numerable, se tiene:

$$\begin{aligned} & f_{X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \\ &= \sum_x f_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) \end{aligned}$$

Para $1 \leq i < j \leq n$, aplicando dos veces la fórmula anterior, se obtiene:

$$\begin{aligned} & f_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \\ &= \sum_{x,y} f_{X_1, \dots, X_n}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n) \end{aligned}$$

Aplicando varias veces la primera fórmula, se obtiene una fórmula similar para la función de densidad conjunta de cualquier subfamilia de las familia de variables aleatorias X_1, \dots, X_n . En particular, se tiene:

$$f_{X_j}(x_j) = \sum_{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n} f_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n)$$

Para el caso de dos variables aleatorias, X, Y , este resultado se escribe de la siguiente manera:

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

En este contexto, las densidades f_X y f_Y son conocidas como **densidades marginales**.

EJEMPLO 1.22. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2x}{N^2(N+1)} & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo. Encuentre las funciones de densidad marginales f_X y f_Y .

Solución

$$f_X(x) = \sum_y f_{X,Y}(x, y) = \frac{2}{N^2(N+1)} \sum_{y=1}^N x = \frac{2x}{N(N+1)}, \text{ para } x \in \{1, \dots, N\}$$

$$f_Y(y) = \sum_x f_{X,Y}(x, y) = \frac{2}{N^2(N+1)} \sum_{x=1}^N x = \frac{1}{N}, \text{ para } y \in \{1, \dots, N\}$$

▲

En el caso absolutamente continuo, se tiene:

$$\begin{aligned} & F_{X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \\ &= \lim_{x \rightarrow \infty} F_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) \\ &= \lim_{x \rightarrow \infty} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{j-1}} \int_{-\infty}^x \int_{-\infty}^{x_{j+1}} \cdots \\ &\quad \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(y_1, \dots, y_{j-1}, y, y_{j+1}, \dots, y_n) dy_n \cdots dy_{j+1} dy_{j-1} \cdots dy_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{j-1}} \int_{-\infty}^{\infty} \int_{-\infty}^{x_{j+1}} \cdots \\ &\quad \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(y_1, \dots, y_{j-1}, y, y_{j+1}, \dots, y_n) dy_n \cdots dy_{j+1} dy_{j-1} \cdots dy_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{j-1}} \int_{-\infty}^{x_{j+1}} \cdots \\ &\quad \cdots \int_{-\infty}^{x_n} \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(y_1, \dots, y_{j-1}, y, y_{j+1}, \dots, y_n) dy dy_n \cdots dy_{j+1} dy_{j-1} \cdots dy_1 \end{aligned}$$

Por lo tanto:

$$= \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) dx$$

es una función de densidad conjunta de $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n$.

Para $1 \leq i < j \leq n$, aplicando dos veces la fórmula anterior, se obtiene:

$$\begin{aligned} & f_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_{j-1}, X_{j+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n) dx dy \end{aligned}$$

Aplicando varias veces la primera fórmula, se obtiene una fórmula similar para una función de densidad conjunta de cualquier subfamilia de las familia de variables aleatorias X_1, \dots, X_n . En particular, se tiene:

$$f_{X_j}(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_n) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_n$$

En particular, para el caso de dos variables aleatorias, X, Y , se tiene:

$$\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \text{ es una función de densidad de } X.$$

$$\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx \text{ es una función de densidad de } Y.$$

Para el caso de tres variables aleatorias, X, Y, Z , se tiene:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dy dz \text{ es una función de densidad de } X.$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dx dy \text{ es una función de densidad de } Y.$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dx dy \text{ es una función de densidad de } Z.$$

$$\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dx \text{ es una función de densidad de } Y, Z.$$

$$\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dy \text{ es una función de densidad de } X, Z.$$

$$\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z) dz \text{ es una función de densidad de } X, Y.$$

Siendo rigurosos, se requiere de una pequeña corrección en lo dicho anteriormente. Por ejemplo, en el caso de un vector aleatorio absolutamente continuo (X, Y) , con función de densidad conjunta $f_{X,Y}$, se tiene:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$$

Así que, de acuerdo con el teorema de Fubini (ver apéndice), el conjunto de puntos $y \in \mathbb{R}$ para los cuales la función $x \mapsto f_{X,Y}(x, y)$ no es integrable, tiene medida cero; es decir, existe un conjunto $A \subset \mathbb{R}$, de medida cero, tal que $\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx < \infty$

para cualquier $y \notin A$. Los puntos $y \in A$ no afectan la integrabilidad de $f_{X,Y}$ como función de dos variables, pero se tiene $\int_{-\infty}^{\infty} f_{X,Y}(x, y)dx = \infty$ para cualquier $y \in A$. Entonces, si bien la función $y \mapsto \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$ es una función de densidad de Y , no toma únicamente valores reales y hasta ahora hemos pedido a toda función de densidad el ser una función real valuada.

Con el objeto de no acarrear este detalle en lo que sigue, cuando tengamos un vector aleatorio absolutamente continuo (X, Y) , con función de densidad conjunta $f_{X,Y}$, tomaremos como funciones de densidad de X y Y a las funciones definidas, respectivamente, como sigue:

$$f_X(x) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy & \text{si } \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy < \infty \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx & \text{si } \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Haremos las mismas consideraciones en el caso de un vector aleatorio absolutamente continuo (X_1, \dots, X_n) .

EJEMPLO 1.23. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{y} e^{-\frac{x}{y}} e^{-y} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad de Y .

Solución

Para $y > 0$, se tiene:

$$f_Y(y) = \int_0^{\infty} \frac{1}{y} e^{-\frac{x}{y}} e^{-y} dx = e^{-y} \int_0^{\infty} \frac{1}{y} e^{-\frac{x}{y}} dx = e^{-y}$$

Así que Y tiene distribución exponencial con parámetro $\lambda = 1$.

EJEMPLO 1.24. Se elige un punto al azar en el interior del círculo de centro en el origen y radio R . Sean X y Y la abscisa y ordenada, respectivamente, del punto seleccionado. Encuentre las funciones de densidad marginales f_X y f_Y .

Solución

Se tiene:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{si } x^2 + y^2 < R^2 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$f_X(x) = \begin{cases} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2} dy & \text{si } x \in (-R, R) \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2-x^2} & \text{si } x \in (-R, R) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} \frac{1}{\pi R^2} dx & \text{si } y \in (-R, R) \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2-y^2} & \text{si } y \in (-R, R) \\ 0 & \text{en otro caso} \end{cases}$$

1.4. Distribuciones conjuntas de variables aleatorias independientes

Recordemos (ver sección 6.4 del primer volumen de este libro) que se dice que n variables aleatorias, X_1, \dots, X_n , son independientes si para cualquier colección de conjuntos borelianos de números reales, A_1, \dots, A_n , los eventos $[X_1 \in A_1], \dots, [X_n \in A_n]$ son independientes.

Además, se tienen los siguientes resultados:

PROPOSICIÓN 1.25. Sean X_1, \dots, X_n n variables aleatorias independientes y f_1, \dots, f_n n funciones borelianas de \mathbb{R} en \mathbb{R} . Entonces las variables aleatorias $f_1(X_1), \dots, f_n(X_n)$ son independientes.

PROPOSICIÓN 1.26. Sean X_1, \dots, X_{n+m} $n+m$ variables aleatorias independientes y $f : \mathbb{R}^n \mapsto \mathbb{R}$ y $g : \mathbb{R}^m \mapsto \mathbb{R}$ dos funciones borelianas cualesquiera. Entonces las variables aleatorias $f(X_1, \dots, X_n)$ y $g(X_{n+1}, \dots, X_{n+m})$ son independientes.

Dadas dos variables aleatorias independientes X y Y , se tiene, para cualquier vector $(x, y) \in \mathbb{R}^2$:

$$F_{X,Y}(x, y) = P[X \leq x, Y \leq y] = P[X \leq x] P[Y \leq y] = F_X(x) F_Y(y)$$

Si, además, tanto X como Y son absolutamente continuas, con funciones de densidad f_X y f_Y , respectivamente, se tiene:

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) = \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv du$$

Por lo tanto, el vector (X, Y) es absolutamente continuo y $f_X f_Y$ es una función de densidad de (X, Y) .

De la misma manera, si tanto X como Y son discretas, con funciones de densidad f_X y f_Y , respectivamente, se tiene:

$$f_{X,Y}(x, y) = P[X = x, Y = y] = P[X = x] P[Y = y] = f_X(x) f_Y(y)$$

A continuación se demuestran los inversos de estos resultados.

PROPOSICIÓN 1.27. *Las variables aleatorias X_1, \dots, X_n son independientes si y sólo si su función de distribución conjunta es igual al producto de las funciones de distribución marginales.*

Demostración

Únicamente se hará la demostración para el caso de dos variables aleatorias. La prueba en el caso n -dimensional es similar.

Sean X y Y dos variables aleatorias tales que $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ para cualquier vector $(x, y) \in \mathbb{R}^2$. Definamos entonces:

$$\mathcal{H} = \left\{ A \subset \mathbb{R} : \begin{array}{l} A \text{ es boreliano y} \\ P[X \in A, Y \leq y] = P[X \in A] P[Y \leq y] \text{ para cualquier } y \in \mathbb{R} \end{array} \right\}$$

\mathcal{H} es entonces un d -sistema que contiene a todos los intervalos de la forma $(-\infty, x]$, los cuales forman un π -sistema que genera a los borelianos. Con base en el teorema de clases monótonas (ver subsección 5.5.1 del primer volumen de este libro), se concluye entonces que \mathcal{H} contiene a todos los borelianos. Es decir, $P[X \in A, Y \leq y] = P[X \in A] P[Y \leq y]$ para cualquier $y \in \mathbb{R}$ y cualquier boreliano $A \subset \mathbb{R}$.

Sea ahora:

$$\mathcal{G} = \left\{ B \subset \mathbb{R} : \begin{array}{l} B \text{ es boreliano y} \\ P[X \in A, Y \in B] = P[X \in A] P[Y \in B] \\ \text{para cualquier boreliano } A \end{array} \right\}$$

Con base en el resultado anterior, \mathcal{G} es un d -sistema que contiene a todos los intervalos de la forma $(-\infty, y]$, los cuales forman un π -sistema que genera a los borelianos. Con base en el teorema de clases monótonas, se concluye entonces que \mathcal{G} contiene a todos los borelianos. Es decir, $P[X \in A, Y \in B] = P[X \in A] P[Y \in B]$ para cualquier pareja de borelianos A y B . Así que X y Y son independientes. ■

PROPOSICIÓN 1.28. *Supongamos que el vector aleatorio (X_1, \dots, X_n) es discreto o absolutamente continuo, entonces, las variables aleatorias que lo forman son independientes si y sólo si el producto de sus funciones de densidad es una función de densidad conjunta de (X_1, \dots, X_n) .*

Demostración

Únicamente se hará la demostración para el caso de dos variables aleatorias. La prueba en el caso n -dimensional es similar.

Sea (X, Y) un vector aleatorio discreto o absolutamente continuo y supongamos $f_X f_Y$ es una función de densidad conjunta de (X, Y) .

En el caso discreto, se tiene:

$$\begin{aligned} F_{X,Y}(x, y) &= \sum_{\{(u,v) \in \mathbb{R}^2 | u \leq x, v \leq y\}} f_{X,Y}(u, v) = \sum_{\{(u,v) \in \mathbb{R}^2 | u \leq x, v \leq y\}} f_X(u) f_Y(v) \\ &= \sum_{\{u \in \mathbb{R} | u \leq x\}} f_X(u) \sum_{\{v \in \mathbb{R} | v \leq y\}} f_Y(v) = F_X(x) F_Y(y) \end{aligned}$$

Mientras que en el caso absolutamente continuo, se tiene:

$$\begin{aligned} F_{X,Y}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv du \\ &= \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv = F_X(x) F_Y(y) \end{aligned}$$

Así que, en cualquier caso, la función de distribución conjunta de X y Y es igual al producto de las funciones de distribución marginales. Por lo tanto, con base en la proposición 1.27, las variables aleatorias X y Y son independientes. ■

Como ejemplo, se puede ver inmediatamente que las variables aleatorias del ejemplo 1.1 son independientes mientras que las del ejemplo 1.2 no lo son.

EJEMPLO 1.29. *Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:*

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{N^2(N+1)}(x + y) & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

¿Son X y Y independientes?

Solución

$$\begin{aligned} f_X(x) &= \sum_y f_{X,Y}(x, y) = \frac{1}{N^2(N+1)} \sum_{y=1}^N (x + y) = \frac{1}{N^2(N+1)} (Nx + \frac{N(N+1)}{2}) \\ &= \frac{x}{N(N+1)} + \frac{1}{2N}, \text{ para } x \in \{1, \dots, N\} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \sum_x f_{X,Y}(x, y) = \frac{1}{N^2(N+1)} \sum_{x=1}^N (x + y) = \frac{1}{N^2(N+1)} (Ny + \frac{N(N+1)}{2}) \\ &= \frac{y}{N(N+1)} + \frac{1}{2N}, \text{ para } y \in \{1, \dots, N\} \end{aligned}$$

Se tiene $f_{X,Y} \neq f_X f_Y$, por lo tanto X y Y no son independientes.

EJEMPLO 1.30. *Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:*

$$f_{X,Y}(x, y) = \begin{cases} \frac{4}{N^2(N+1)^2}xy & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

¿Son X y Y independientes?

Solución

$$f_X(x) = \sum_y f_{X,Y}(x, y) = \frac{4}{N^2(N+1)^2} \sum_{y=1}^N xy = \frac{2x}{N(N+1)}, \text{ para } x \in \{1, \dots, N\}$$

$$f_Y(y) = \sum_x f_{X,Y}(x, y) = \frac{4}{N^2(N+1)^2} \sum_{x=1}^N xy = \frac{2y}{N(N+1)}, \text{ para } y \in \{1, \dots, N\}$$

Se tiene $f_{X,Y} = f_X f_Y$, por lo tanto X y Y son independientes.

EJEMPLO 1.31. Sea Y_1, Y_2, \dots una sucesión de variables aleatorias independientes, todas ellas con distribución Bernoulli de parámetro p y, para $k \in \mathbb{N}$, definamos:

$$T_k = \inf \left\{ j \in \mathbb{N} : \sum_{i=1}^j Y_i = k \right\}$$

Si vemos a las variables aleatorias Y_1, Y_2, \dots como una sucesión de ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es p , las variables aleatorias T_1, T_2, \dots son los tiempos en que ocurre éxito. Vamos a encontrar, para cada $n \in \mathbb{N}$, la función de densidad conjunta de $Y_1 = T_1, Y_2 = T_2 - T_1, Y_3 = T_3 - T_2, \dots, Y_n = T_n - T_{n-1}$.

Definamos $X_0 = 0$ y, para $k \in \mathbb{N}$, $X_k = \sum_{i=1}^k Y_i$, entonces:

- (i) Si $0 < k_1 < \dots < k_n$, entonces las variables aleatorias $X_{k_1}, X_{k_2} - X_{k_1}, \dots, X_{k_n} - X_{k_{n-1}}$ son independientes.
- (ii) Si $j < k$, entonces la variable aleatoria $X_k - X_j$ tiene distribución binomial con parámetros $n = k - j$ y p .

Para $y_1, \dots, y_n \in \mathbb{N}$, se tiene:

$$\begin{aligned} & P [T_1 = y_1, T_2 - T_1 = y_2, \dots, T_n - T_{n-1} = y_n] \\ &= P [T_1 = y_1, T_2 = y_1 + y_2, \dots, T_n = y_1 + y_2 + \dots + y_n] \\ &= P [X_{y_1-1} = 0, X_{y_1} = 1, X_{y_1+y_2-1} = 1, X_{y_1+y_2} = 2, \dots, \\ & \quad X_{y_1+\dots+y_{n-1}} = n-1, X_{y_1+\dots+y_n} = n] \\ &= P [X_{y_1-1} = 0, X_{y_1} - X_{y_1-1} = 1, X_{y_1+y_2-1} - X_{y_1} = 0, X_{y_1+y_2} - X_{y_1+y_2-1} = 1, \dots, \\ & \quad X_{y_1+\dots+y_{n-1}} - X_{y_1+\dots+y_{n-1}} = 0, X_{y_1+\dots+y_n} - X_{y_1+\dots+y_{n-1}} = 1] \end{aligned}$$

$$= (1-p)^{y_1-1} p (1-p)^{y_2-1} p \cdots (1-p)^{y_n-1} p$$

Así que $T_1, T_2 - T_1, \dots, T_n - T_{n-1}$ son independientes y todas ellas tienen distribución geométrica con parámetro p .

EJEMPLO 1.32. Supongamos ahora que un cierto evento ocurre en los tiempos aleatorios, T_1, T_2, \dots , de tal manera que las variables aleatorias $Y_1 = T_1, Y_2 = T_2 - T_1, Y_3 = T_3 - T_2, \dots$ son independientes y todas ellas tienen distribución geométrica con parámetro p . Vamos a demostrar que existe una sucesión de variables aleatorias independientes, Y_1, Y_2, \dots , todas ellas con distribución Bernoulli de parámetro p tales que, para $k \in \mathbb{N}$, $T_k = \inf \left\{ j \in \mathbb{N} : \sum_{i=1}^j Y_i = k \right\}$.

Para cada $k \in \{0, 1, \dots\}$, sea X_k el número de veces que ocurre el evento hasta el tiempo n y definamos $Y_1 = X_1, Y_2 = X_2 - X_1, Y_3 = X_3 - X_2, \dots$. Entonces, evidentemente, se tiene, para $k \in \mathbb{N}$, $T_k = \inf \left\{ j \in \mathbb{N} : \sum_{i=1}^j Y_i = k \right\}$.

Sean $k_1, k_2, \dots, k_n \in \{0, 1\}$, $r = k_1 + k_2 + \dots + k_n$ y $k_{j_1} = \dots = k_{j_r} = 1$, con $j_1, \dots, j_r \in \{1, \dots, n\}$, $j_1 < \dots < j_r$. Entonces:

$$\begin{aligned} & P[Y_1 = k_1, Y_2 = k_2, \dots, Y_n = k_n] \\ &= P[X_1 = k_1, X_2 - X_1 = k_2, \dots, X_n - X_{n-1} = k_n] \\ &= P[X_1 = k_1, X_2 = k_1 + k_2, \dots, X_n = k_1 + k_2 + \dots + k_n] \\ &= P[T_1 = j_1, \dots, T_r = j_r, T_{r+1} > n] \\ &= P[T_1 = j_1, T_2 - T_1 = j_2 - j_1, \dots, T_r - T_{r-1} = j_r - j_{r-1}, T_{r+1} - T_r > n - j_r] \\ &= p(1-p)^{j_1-1} p(1-p)^{j_2-j_1-1} \cdots p(1-p)^{j_r-j_{r-1}-1} (1-p)^{n-j_r} \\ &= p^r (1-p)^{n-r} \\ &= p^{k_1+k_2+\dots+k_n} (1-p)^{n-(k_1+k_2+\dots+k_n)} \\ &= p^{k_1} (1-p)^{1-k_1} p^{k_2} (1-p)^{1-k_2} \cdots p^{k_n} (1-p)^{1-k_n} \end{aligned}$$

Así que las variables aleatorias Y_1, \dots, Y_n son independientes y todas ellas tienen distribución Bernoulli con parámetro p .

▲

Cuando se tiene una sucesión de ensayos de Bernoulli independientes en cada uno de los cuales la probabilidad de éxito es p , se puede pensar esta sucesión como una sucesión de eventos que ocurren aleatoriamente en tiempos enteros positivos de tal

manera que, para cada entero k , la probabilidad de que ocurra el evento en el tiempo k es igual a p . Si definimos X_k como el número de veces que ocurre el evento hasta el tiempo k , entonces el proceso X_k tiene las siguientes propiedades:

- (i) $X_0 = 0$
- (ii) Si $0 < k_1 < \dots < k_n$, entonces las variables aleatorias $X_{k_1}, X_{k_2} - X_{k_1}, \dots, X_{k_n} - X_{k_{n-1}}$ son independientes.
- (iii) Si $j < k$, entonces la variable aleatoria $X_k - X_j$ tiene distribución binomial con parámetros $n = k - j$ y p .

Inversamente, si un proceso X_k satisface las condiciones *i*, *ii* y *iii*, entonces la sucesión $X_1, X_2 - X_1, X_3 - X_2, \dots$ constituye una sucesión de ensayos de Bernoulli independientes en cada uno de los cuales la probabilidad de éxito es p . Además, como $X_k = \sum_{j=1}^k X_j - X_{j-1}$, X_k es igual al número de éxitos hasta el tiempo k .

EJEMPLO 1.33. *Consideremos una sucesión de ensayos de Bernoulli independientes en cada uno de los cuales la probabilidad de éxito es igual a p y, para $k \in \mathbb{N}$, sea X_k el número de fracasos antes del k -ésimo éxito. Vamos a encontrar, para cada $n \in \mathbb{N}$, la función de densidad conjunta de $X_1, X_2 - X_1, \dots, X_n - X_{n-1}$.*

Para $j_1, \dots, j_n \in \mathbb{N}$ tales que $j_1 < \dots < j_n$, se tiene:

$$P[X_1 = j_1, \dots, X_n = j_n] = (1-p)^{j_1} p (1-p)^{j_2 - j_1} p \dots (1-p)^{j_n - j_{n-1}} p = (1-p)^{j_n} p^n$$

Así que, para $k_1, \dots, k_n \in \mathbb{N}$, se tiene:

$$\begin{aligned} P[X_1 = k_1, X_2 - X_1 = k_2, \dots, X_n - X_{n-1} = k_n] \\ = P[X_1 = k_1, X_2 = k_1 + k_2, \dots, X_n = k_1 + k_2 + \dots + k_n] \\ = (1-p)^{k_1 + k_2 + \dots + k_n} p^n = (1-p)^{k_1} p (1-p)^{k_2} p \dots (1-p)^{k_n} p \end{aligned}$$

Así que $X_1, X_2 - X_1, \dots, X_n - X_{n-1}$ son independientes y todas tienen distribución geométrica con parámetro p .

▲

En el caso de dos variables aleatorias independientes, X, Y , que sean discretas o absolutamente continuas, el cálculo de una probabilidad $P[(X, Y) \in A]$, para un subconjunto $A \subset \mathbb{R}^2$, se simplifica gracias a la proposición 1.28 y porque una probabilidad de la forma $P[X \in B, Y \in C]$ puede obtenerse como el producto $P[X \in B] P[Y \in C]$.

EJEMPLO 1.34. *Sean X y Y variables aleatorias independientes, cada una de las cuales tiene distribución geométrica con parámetro p . Encuentre a) $P(X > 2Y)$ y b) $P(Y \geq X)$.*

Solución

$$\begin{aligned}
a. P[X > 2Y] &= \sum_{y=0}^{\infty} P[X > 2Y, Y = y] = \sum_{y=0}^{\infty} P[X > 2y, Y = y] \\
&= \sum_{y=0}^{\infty} P[X > 2y] P[Y = y] = \sum_{y=0}^{\infty} (1-p)^{2y+1} p(1-p)^y \\
&= p(1-p) \sum_{y=0}^{\infty} (1-p)^{3y} = p(1-p) \frac{1}{[1-(1-p)^3]} = \frac{p(1-p)}{1-(1-p)^3} \\
b. P[Y \geq X] &= \sum_{x=0}^{\infty} P[Y \geq X, X = x] = \sum_{x=0}^{\infty} P[Y \geq x, X = x] \\
&= \sum_{x=0}^{\infty} P[Y \geq x] P[X = x] = \sum_{x=0}^{\infty} (1-p)^x p(1-p)^x \\
&= p \sum_{x=0}^{\infty} (1-p)^{2x} = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}
\end{aligned}$$

EJEMPLO 1.35. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial, X de parámetro λ_1 y Y de parámetro λ_2 . Encuentre $P[X < Y]$.

Solución

$$\begin{aligned}
P[X < Y] &= \int_0^{\infty} \int_x^{\infty} \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y} dy dx = \int_0^{\infty} \lambda_1 e_2^{-\lambda_1 x} e^{-\lambda_2 x} dx \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2} \int_0^{\infty} (\lambda_1 + \lambda_2) e_2^{-(\lambda_1 + \lambda_2)x} dx = \frac{\lambda_1}{\lambda_1 + \lambda_2}
\end{aligned}$$

EJEMPLO 1.36. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre a) $P[X < 2Y]$ y b) $P[2X < 3Y + 1]$.

Solución

$$\begin{aligned}
a. P[X < 2Y] &= \int_0^{\infty} \int_{\frac{x}{2}}^{\infty} \lambda^2 e^{-\lambda x} e^{-\lambda y} dy dx = \int_0^{\infty} \lambda e^{-\frac{3}{2}x} dx = \frac{2}{3} \\
b. P[2X < 3Y + 1] &= \int_0^{\infty} \int_0^{\frac{1}{2}(3y+1)} \lambda^2 e^{-\lambda x} e^{-\lambda y} dx dy \\
&= \int_0^{\infty} \lambda e^{-\lambda y} \left[1 - e^{-\frac{\lambda}{2}(3y+1)} \right] dy = \int_0^{\infty} \lambda e^{-\lambda y} dy - \int_0^{\infty} \lambda e^{-\frac{\lambda}{2}(5y+1)} dy = 1 - \frac{2}{5} e^{-\frac{1}{2}\lambda}
\end{aligned}$$

EJEMPLO 1.37. Sean X , Y y Z tres variables aleatorias independientes, las tres con distribución uniforme en el intervalo $(0, 1)$. Encuentre $P[X + Y > \frac{3}{2}Z]$.

Solución

$$\begin{aligned}
P[X + Y > \frac{3}{2}Z] &= \int_0^{\frac{2}{3}} \left(\int_0^{\frac{3z}{2}} \int_{-x+\frac{3z}{2}}^1 dy dx + \int_{\frac{3z}{2}}^1 \int_0^1 dy dx \right) dz \\
&+ \int_{\frac{2}{3}}^1 \int_{\frac{3z}{2}-1}^1 \int_{-x+\frac{3z}{2}}^1 dy dx dz = \frac{5}{9} + \frac{7}{72} = \frac{47}{72}
\end{aligned}$$

$$z \in \left(0, \frac{2}{3}\right)$$

$$z \in \left[\frac{2}{3}, 1\right)$$

EJERCICIOS

EJERCICIO 1.1. *Para cada una de las siguientes funciones, determine si F es una función de distribución conjunta y, de ser así, diseñe un experimento aleatorio y defina dos variables aleatorias cuya función de distribución conjunta sea F .*

$$a) F(x, y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0 \\ y & \text{si } x \geq 0, 0 \leq y < 1 \\ 1 & \text{si } x \geq 0, y \geq 1 \end{cases}$$

$$b) F(x, y) = \begin{cases} 0 & \text{si } x < 1 \text{ ó } y < 0 \\ x^2y & \text{si } x^2y < 1, x \geq 1, y \geq 0 \\ 1 & \text{si } x^2y \geq 1, x \geq 1, y \geq 0 \end{cases}$$

EJERCICIO 1.2. *Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del rombo de vértices $A(2, 0)$, $B(0, 1)$, $C(-2, 0)$ y $D(0, -1)$. Sean X y Y la abscisa y la ordenada, respectivamente, del punto seleccionado. Encuentre los valores de la función de distribución conjunta, $F_{X,Y}(x, y)$, de X y Y , para las parejas (x, y) pertenecientes al primer cuadrante.*

EJERCICIO 1.3. *Demuestre la proposición 1.6.*

EJERCICIO 1.4. *Consideremos el experimento aleatorio consistente en lanzar 10 veces un par de dados y definamos X como el número de veces en que no se obtiene 5 en ninguno de los dos dados y Y como el número de veces en que se obtiene 5 en los dos dados. Encuentre la función de densidad conjunta de X y Y .*

EJERCICIO 1.5. *Un experimento aleatorio admite únicamente 3 posibles resultados, e_1 , e_2 y e_3 , con probabilidades p_1 , p_2 y $p_3 = 1 - p_1 - p_2$, respectivamente. Supongamos que este experimento se repite, en forma independiente, n veces y, para $i \in \{1, 2, 3\}$, llamemos X_i al número de veces que ocurre e_i . Encuentre a) la función de densidad de $X_1 + X_2$ y b) para $z \in \{0, \dots, n\}$ y $y \in \{0, \dots, z\}$, $P[X_2 = y \mid X_1 + X_2 = z]$.*

EJERCICIO 1.6. *$2r$ bolas se van colocando una por una al azar en cualquiera de r cajas. Sea X_i el número de bolas que quedan en la caja i . Encuentre la función de densidad conjunta de X_1, \dots, X_r .*

EJERCICIO 1.7. *Cada una de N partículas se coloca al azar en una de M celdas. Supongamos que N tiene distribución Poisson con parámetro λ y, para $k \in \{1, \dots, M\}$, llamemos X_k al número de partículas que quedan colocadas en la celda número k . Demuestre que las variables aleatorias X_1, \dots, X_M son independientes y que cada una de ellas tiene distribución Poisson.*

EJERCICIO 1.8 (**Distribución hipergeométrica multivariada**). *Una urna contiene m_1 bolas rojas, m_2 bolas blancas y m_3 bolas negras. Se eligen, al azar y sin reemplazo, n bolas de la urna y se definen las variables aleatorias X_1 , X_2 y X_3 como el número de bolas rojas, blancas y negras, respectivamente, que se obtienen en la muestra. Encuentre la función de densidad conjunta de X_1, X_2, X_3 .*

EJERCICIO 1.9. *Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:*

$$f_{X,Y}(x, y) = \begin{cases} cx & \text{si } x, y \in \{1, \dots, N^2\}, x \leq y^2 \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un número natural y c es una constante. Encuentre a) $P[X = Y]$, b) $P[X < Y]$ y c) $P[X > Y]$.

EJERCICIO 1.10. *Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:*

$$f(x, y) = \begin{cases} \frac{1}{y} e^{-\frac{x}{y}} e^{-y} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $P[X < Y]$.

EJERCICIO 1.11. Sea X una variable aleatoria con distribución normal estándar. Encuentre la función de distribución conjunta de X y $Y = X^2$. ¿Existe una función de densidad conjunta de X y Y ? Justifique su respuesta.

EJERCICIO 1.12. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$ y sea $Z = \min(X, Y)$. ¿Existe una función de densidad conjunta de Z y Y ? Justifique su respuesta.

EJERCICIO 1.13. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} cx & \text{si } x, y \in \{1, \dots, N^2\}, x \leq y^2 \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un número natural y c es una constante. Encuentre las funciones de densidad marginales f_X y f_Y .

EJERCICIO 1.14. Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del triángulo de vértices $A(0, 0)$, $B(1, 0)$ y $C(0, 1)$. Sean X y Y la abscisa y la ordenada, respectivamente, del punto seleccionado. Encuentre a) la función de distribución conjunta de X y Y y b) las densidades marginales f_X y f_Y .

EJERCICIO 1.15. Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del rombo de vértices $A(2, 0)$, $B(0, 1)$, $C(-2, 0)$ y $D(0, -1)$. Sean X y Y la abscisa y la ordenada, respectivamente, del punto seleccionado. Encuentre una función de densidad conjunta de X y Y y las densidades marginales f_X y f_Y .

EJERCICIO 1.16. Sea (X, Y) un vector aleatorio con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{2}xy & \text{si } 0 < y < x < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre las funciones de densidad marginales f_X y f_Y .

EJERCICIO 1.17. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{si } 0 \leq x \leq y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $P[2X < Y]$ y b) las funciones de densidad de X y Y .

EJERCICIO 1.18. *Un sistema con dos estados, 0 y 1, funciona cambiando de estado, en cada unidad de tiempo, con probabilidad $\frac{2}{3}$. Sea X_i el estado del sistema en el tiempo i y supongamos $X_1 = 0$. Para $n \in \{2, 3, \dots\}$, encuentre la función de densidad conjunta de X_1, \dots, X_n .*

EJERCICIO 1.19. *Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:*

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{N^2(N+1)(2N+1)}x^2 & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

¿Son X y Y independientes?

EJERCICIO 1.20. *Consideremos una sucesión de ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es igual a p , y, para $k \in \mathbb{N}$, sea X_k el número de ensayo en el cual ocurre el k -ésimo éxito. Encuentre la probabilidad de que haya n fracasos antes del primer éxito dado que hay $n + m$ fracasos antes del segundo éxito.*

EJERCICIO 1.21. *Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre a) $P[Y \geq 3X - 1]$, b) $P[|X - Y| = 1]$ y c) $P[|X - Y| \leq 2]$.*

EJERCICIO 1.22. *Sean X y Y dos variables aleatorias independientes, X con distribución geométrica de parámetro p , Y con distribución Poisson de parámetro λ . Encuentre $P[Y > X]$.*

EJERCICIO 1.23. *Cada una de dos personas lanza una moneda n veces. Encuentre la probabilidad de que obtengan el mismo número de caras.*

EJERCICIO 1.24. *Sean X y Y dos variables aleatorias independientes, cada una distribuida uniformemente en el conjunto $\{0, \dots, N\}$. Encuentre a) $P(X \geq Y)$ y b) $P(X = Y)$.*

EJERCICIO 1.25. *Sean X y Y dos variables aleatorias independientes, ambas distribuidas uniformemente en el conjunto $\{1, \dots, 2N\}$. Encuentre $P(Y > 2X)$.*

EJERCICIO 1.26. *Sea X una variable aleatoria distribuida uniformemente en el conjunto $\{1, \dots, N\}$ y sea Y una variable aleatoria con distribución geométrica de parámetro p . Suponiendo que X y Y son independientes, encuentre a) $P[|Y - X| = 2]$ y b) $P[Y \geq X]$.*

EJERCICIO 1.27. *Sea X una variable aleatoria con distribución geométrica de parámetro $p_1 = \frac{1}{4}$ y sea Y una variable aleatoria con distribución geométrica de parámetro $p_2 = \frac{1}{3}$. Suponiendo que X y Y son independientes, encuentre $P[Y \geq 2X + 1]$.*

EJERCICIO 1.28. Sea X una variable aleatoria con distribución Poisson de parámetro λ y sea Y una variable aleatoria con distribución geométrica de parámetro p . Suponiendo que X y Y son independientes, encuentre a) $P[Y = X + 2]$ y b) $P[Y \geq X]$.

EJERCICIO 1.29. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal de parámetros $\mu_X = 6, \sigma_X^2 = 1$ y $\mu_Y = 7, \sigma_Y^2 = 2$, respectivamente. Encuentre $P[X > Y]$.

EJERCICIO 1.30. Los tiempos que les toma a dos estudiantes resolver un problema son independientes y ambos tienen distribución exponencial con parámetro λ . Encuentre la probabilidad de que el primer estudiante requiera por lo menos del doble de tiempo que el segundo para resolver un problema.

EJERCICIO 1.31. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre $P[X + Y \leq 1, X - Y \leq 1]$.

EJERCICIO 1.32. Se seleccionan, al azar y de manera independiente, 3 puntos sobre el segmento $[0, L]$. Si X_1, X_2, X_3 son los puntos seleccionados, ¿cuál es la probabilidad de que X_2 quede comprendido entre X_1 y X_3 ?

EJERCICIO 1.33. Sean X, Y y Z tres variables aleatorias independientes, las 3 con distribución exponencial de parámetro λ . Encuentre $P[X + Y < 2Z]$.

EJERCICIO 1.34. Sean X_1, X_2 y X_3 tres variables aleatorias independientes, las 3 con distribución exponencial, de parámetros λ_1, λ_2 y λ_3 , respectivamente. Encuentre $P[X_1 + X_2 < 2X_3]$.

EJERCICIO 1.35. Sean X, Y y Z tres variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $P[X + Y < 2Z]$, b) $P[X + Y > 3Z]$ y c) $P[2Y - X < 2Z]$.

EJERCICIO 1.36. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros $\alpha = 2$ y λ . Encuentre $P[-1 \leq Y - 2X \leq 3]$.

EJERCICIO 1.37. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentre $P[1 \leq X^2 + Y^2 \leq 2]$.

EJERCICIO 1.38. Sean X, Y y Z tres variables aleatorias independientes, X y Y con distribución exponencial de parámetro λ y Z con distribución uniforme en el intervalo $(0, 1)$. Encuentre $P[X + Y < Z]$.

EJERCICIO 1.39. Se seleccionan, al azar y de manera independiente, dos puntos, X y Y , sobre los segmentos $[0, L]$ y $[L, 2L]$, respectivamente. Encuentre la probabilidad de que la distancia entre los dos puntos seleccionados sea mayor que $\frac{1}{3}L$.

EJERCICIO 1.40. Sean X y Y dos variables aleatorias continuas con función de densidad conjunta f . Encuentre una función de densidad conjunta de X^2 y Y^2 . Muestre además que si X y Y son independientes entonces X^2 y Y^2 también lo son.

EJERCICIO 1.41. *En una parada de autobús, el tiempo de llegada de éste se distribuye uniformemente en el intervalo que va de las 7 : 00 a las 7 : 15 hrs. y el siguiente autobús pasa exactamente 15 minutos después del primero. Si el tiempo de llegada de una persona a esa parada se distribuye uniformemente en el intervalo que va de las 7 : 10 a las 7 : 15 hrs., a) ¿cuál es la probabilidad de que la persona llegue a la parada antes que el primer autobús?, b) si T es el tiempo que espera la persona, desde que llega a la parada hasta que pasa un autobús, encuentre e identifique la distribución de T .*

EJERCICIO 1.42. *Tres números, a , b y c , se eligen al azar y de manera independiente en el intervalo $(0, 1)$. ¿Cuál es la probabilidad de que las raíces de la ecuación $ax^2 + bx + c = 0$ sean a) reales? y b) iguales?*

CAPÍTULO 2

DISTRIBUCIONES DE FUNCIONES DE VECTORES ALEATORIOS

Cuando una suspensión de pequeñas partículas en un líquido es vista bajo el microscopio, las partículas parecen animadas con un peculiar movimiento azaroso - el movimiento browniano. Este movimiento es de una naturaleza tan irregular que Perrin dice de él: "Uno se da cuenta con tales ejemplos que tan cerca están los matemáticos de la verdad al rechazar, por un instinto lógico, las pretendidas demostraciones geométricas, las cuales son vistas como evidencia experimental de la existencia de una tangente en cada punto de una curva". De aquí que se convierta en un tema de interés para los matemáticos el descubrir cuáles son las condiciones que definen a las trayectorias de estas partículas.

Norbert Wiener

2.1. Distribuciones de funciones de vectores aleatorios discretos

En esta sección vamos a utilizar la relación 1.1 del capítulo anterior para encontrar la función de densidad de una función de un vector aleatorio discreto.

EJEMPLO 2.1. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2x}{N^2(N+1)} & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo. Encuentre la función de densidad de a) $U = X + Y$,
b) $V = Y - X$.

Solución

$$\begin{aligned}
a. f_U(u) &= P[X + Y = u] = \sum_{x=1}^N P[X + Y = u, X = x] \\
&= \sum_{x=1}^N P[x + Y = u, X = x] = \sum_{x=1}^N P[X = x, Y = u - x] \\
&= \begin{cases} \sum_{x=1}^{u-1} \frac{2x}{N^2(N+1)} & \text{si } u \in \{2, \dots, N+1\} \\ \sum_{x=u-N}^N \frac{2x}{N^2(N+1)} & \text{si } u \in \{N+2, \dots, 2N\} \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} \frac{u(u-1)}{N^2(N+1)} & \text{si } u \in \{2, \dots, N+1\} \\ \frac{u(2N+1-u)}{N^2(N+1)} & \text{si } u \in \{N+2, \dots, 2N\} \\ 0 & \text{en otro caso} \end{cases}
\end{aligned}$$

$$\begin{aligned}
b. f_V(v) &= P[Y - X = v] = \sum_{x=1}^N P[Y - X = v, X = x] \\
&= \sum_{x=1}^N P[Y - x = v, X = x] = \sum_{x=1}^N P[X = x, Y = x + v] \\
&= \begin{cases} \sum_{x=1-v}^N \frac{2x}{N^2(N+1)} & \text{si } v \in \{1-N, \dots, -1\} \\ \sum_{x=1}^{N-v} \frac{2x}{N^2(N+1)} & \text{si } v \in \{0, \dots, N-1\} \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} \frac{(N+v)(N-v+1)}{N^2(N+1)} & \text{si } v \in \{1-N, \dots, -1\} \\ \frac{(N-v)(N-v+1)}{N^2(N+1)} & \text{si } v \in \{0, \dots, N-1\} \\ 0 & \text{en otro caso} \end{cases}
\end{aligned}$$

EJEMPLO 2.2. Sean X y Y variables aleatorias independientes, cada una de las cuales tiene distribución geométrica con parámetro p . Encuentre:

- la función de densidad de $\min(X, Y)$.
- la función de densidad de $X + Y$.
- $P(Y = y | X + Y = z)$ para $y \in \{0, \dots, z\}$.

Solución

a. Sea $Z = \min(X, Y)$, entonces, para $x \in \{0, 1, \dots\}$, se tiene:

$$P(Z \geq x) = P(X \geq x, Y \geq x) = P(X \geq x)P(Y \geq x) = (1-p)^x(1-p)^x = (1-p)^{2x}$$

Por lo tanto, Z tiene distribución geométrica con parámetro $1 - (1-p)^2 = p(2-p)$.

b. Los posibles valores de $X + Y$ son $z = 0, 1, \dots$. Para un valor z de éstos, se tiene:

$$\begin{aligned} P[X + Y = z] &= \sum_{x=0}^{\infty} P[X + Y = z, X = x] = \sum_{x=0}^{\infty} P[Y = z - x, X = x] \\ &= \sum_{x=0}^z P[Y = z - x] P[X = x] = \sum_{x=0}^z p(1-p)^{z-x} p(1-p)^x = \sum_{x=0}^z p^2(1-p)^z \\ &= (z+1)p^2(1-p)^z \end{aligned}$$

Por lo tanto, $X + Y$ tiene distribución binomial negativa con parámetros p y $r = 2$.

$$\begin{aligned} c. P[Y = y | X + Y = z] &= \frac{P[Y=y, X+Y=z]}{P[X+Y=z]} = \frac{P[Y=y, X=z-y]}{P[X+Y=z]} = \frac{P[Y=y]P[X=z-y]}{P[X+Y=z]} \\ &= \frac{p(1-p)^y p(1-p)^{z-y}}{(z+1)p^2(1-p)^z} = \frac{p^2(1-p)^z}{(z+1)p^2(1-p)^z} = \frac{1}{z+1} \end{aligned}$$

Es decir, dado que $X + Y = z$, Y tiene distribución uniforme en el conjunto $\{0, \dots, z\}$.

EJEMPLO 2.3. Sean X y Y variables aleatorias independientes, ambas con distribución binomial negativa de parámetros (r, p) y (s, p) , respectivamente. Demuestre que $X + Y$ tiene distribución binomial negativa con parámetros $(r + s, p)$.

Solución

Los posibles valores de $X + Y$ son $z = 0, 1, \dots$. Para un valor z de éstos, se tiene:

$$\begin{aligned} P[X + Y = z] &= \sum_{x=0}^{\infty} P[X + Y = z, X = x] = \sum_{x=0}^{\infty} P[Y = z - x, X = x] \\ &= \sum_{x=0}^z P[Y = z - x] P[X = x] = \sum_{x=0}^z \binom{s+z-x-1}{z-x} p^s (1-p)^{z-x} \binom{r+x-1}{x} p^r (1-p)^x \\ &= p^{r+s} (1-p)^z \sum_{x=0}^z \binom{s+z-x-1}{z-x} \binom{r+x-1}{x} \end{aligned}$$

Pero, de acuerdo con el lema 2.4, el cual se enuncia y demuestra al concluir este ejemplo, se tiene:

$$\sum_{x=0}^z \binom{s+z-x-1}{z-x} \binom{r+x-1}{x} = \binom{r+s+z-1}{z}$$

Así que:

$$P[X + Y = z] = \binom{r+s+z-1}{z} p^{r+s} (1-p)^z$$

LEMA 2.4. Para $r, s \in \mathbb{N}$ y $z \geq 0$, se tiene: $\sum_{k=0}^z \binom{r+k-1}{k} \binom{s+z-k-1}{z-k} = \binom{r+s+z-1}{z}$.

Demostración

Para $-1 < t < 1$ y $r, s \in \mathbb{N}$, se tiene:

$$(1-t)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} t^k$$

$$(1-t)^{-s} = \sum_{k=0}^{\infty} \binom{s+k-1}{k} t^k$$

$$(1-t)^{-(r+s)} = \sum_{k=0}^{\infty} \binom{r+s+k-1}{k} t^k$$

Por lo tanto:

$$\sum_{k=0}^{\infty} \binom{r+k-1}{k} t^k \sum_{k=0}^{\infty} \binom{s+k-1}{k} t^k = \sum_{k=0}^{\infty} \binom{r+s+k-1}{k} t^k$$

Igualando los coeficientes de t^z , se obtiene el resultado. ■

2.2. Distribuciones de funciones de vectores aleatorios continuos

En esta sección abordaremos el problema de encontrar la función de densidad de una función de un vector aleatorio absolutamente continuo. La propiedad básica que utilizaremos en este caso es la relación 1.2.

EJEMPLO 2.5. *Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Encuentre fórmulas para las funciones de densidad de las variables aleatorias $U = X + Y$ y $V = Y - X$.*

Solución

$$\begin{aligned} F_U(u) &= P[X + Y \leq u] = \iint_{\{(x,y) \in \mathbb{R}^2: x+y \leq u\}} f_{X,Y}(x,y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{u-x} f_{X,Y}(x,y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^u f_{X,Y}(x, z-x) dz dx \\ &= \int_{-\infty}^u \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) dx du \end{aligned}$$

Por lo tanto:

$$\int_{-\infty}^{\infty} f_{X,Y}(x, u-x) dx$$

es una función de densidad de U .

$$\begin{aligned} F_V(v) &= P[Y - X \leq v] = \iint_{\{(x,y) \in \mathbb{R}^2: y-x \leq v\}} f_{X,Y}(x,y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{u+x} f_{X,Y}(x,y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^u f_{X,Y}(x, z+x) dz dx \\ &= \int_{-\infty}^u \int_{-\infty}^{\infty} f_{X,Y}(x, z+x) dx du \end{aligned}$$

Por lo tanto:

$$\int_{-\infty}^{\infty} f_{X,Y}(x, v+x) dx$$

es una función de densidad de V .

COMENTARIO 2.6. La función $z \mapsto \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$ es llamada la convolución de f_X y f_Y y se denota por $f_X * f_Y$. Así que, si X y Y son independientes, $f_X * f_Y$ es una función de densidad de $X + Y$. Además:

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x,y)dydx = \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)dx$$

Así que $F_{X+Y} = f_X * F_Y$.

EJEMPLO 2.7. Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} \frac{2}{11}(4-x-y) & \text{si } 0 < x < 1, 0 < y < x+2 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad de a) $U = X + Y$ y b) $V = Y - X$.

Solución

a. Se tiene $f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx$.

Ahora bien, para que podamos aplicar la fórmula $f_{X,Y}(x,y) = \frac{2}{11}(4-x-y)$ dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx$, se requiere tener $0 < x < 1$ y $0 < u-x < x+2$, es decir $0 < x < 1$ y $\frac{u-2}{2} < x < u$. Esta región se representa en la siguiente figura:

En la figura puede verse que, con el objeto de especificar el rango de valores de x dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx$, conviene partir el rango de valores de u en tres intervalos, a saber $(0, 1)$, $[1, 2)$ y $[2, 4)$, obteniéndose entonces:

$$f_U(u) = \begin{cases} \int_0^u f_{X,Y}(x, u-x)dx & \text{si } 0 < u < 1 \\ \int_0^1 f_{X,Y}(x, u-x)dx & \text{si } 1 \leq u < 2 \\ \int_{\frac{u-2}{2}}^1 f_{X,Y}(x, u-x)dx & \text{si } 2 \leq u < 4 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{2}{11} \int_0^u (4-u)dx & \text{si } 0 < u < 1 \\ \frac{2}{11} \int_0^1 (4-u)dx & \text{si } 1 \leq u < 2 \\ \frac{2}{11} \int_{\frac{u-2}{2}}^1 (4-u)dx & \text{si } 2 \leq u < 4 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{2}{11}u(4-u) & \text{si } 0 < u < 1 \\ \frac{2}{11}(4-u) & \text{si } 1 \leq u < 2 \\ \frac{1}{11}(4-u)^2 & \text{si } 2 \leq u < 4 \\ 0 & \text{en otro caso} \end{cases}$$

b. Se tiene $f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, v+x)dx$.

Ahora bien, para que podamos aplicar la fórmula $f_{X,Y}(x, y) = \frac{2}{11}(4-x-y)$ dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, v+x)dx$, se requiere tener $0 < x < 1$ y $0 < v+x < x+2$, es decir $0 < x < 1$, $x > -v$ y $v < 2$. Esta región se representa en la siguiente figura:

En la figura puede verse que, con el objeto de especificar el rango de valores de x dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, v+x)dx$, conviene partir el rango de valores de v en dos intervalos, a saber $(-1, 0)$ y $[0, 2)$, obteniéndose:

$$f_V(v) = \begin{cases} \int_{-v}^1 f_{X,Y}(x, v+x)dx & \text{si } -1 < v < 0 \\ \int_0^1 f_{X,Y}(x, v+x)dx & \text{si } 0 \leq v < 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{2}{11} \int_{-v}^1 (4-v-2x)dx & \text{si } -1 < v < 0 \\ \frac{2}{11} \int_0^1 (4-v-2x)dx & \text{si } 0 \leq v < 2 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{6}{11}(1+v) & \text{si } -1 < v < 0 \\ \frac{2}{11}(3-v) & \text{si } 0 \leq v < 2 \\ 0 & \text{en otro caso} \end{cases}$$

EJEMPLO 2.8. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad de a) $U = X + Y$ y b) $V = Y - X$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{si } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

a. Se tiene $f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u - x)dx$.

Ahora bien, para que podamos sustituir $f_{X,Y}(x, u - x)$ por 1 dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, u - x)dx$, se requiere tener $0 < x < 1$ y $0 < u - x < 1$, es decir $0 < x < 1$ y $u - 1 < x < u$. Esta región se representa en la siguiente figura:

En la figura puede verse que, con el objeto de especificar el rango de valores de x dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, u - x)dx$, conviene partir el rango de valores de u en dos intervalos, a saber $(0, 1)$ y $[1, 2)$, obteniéndose entonces:

$$f_U(u) = \begin{cases} \int_0^u f_{X,Y}(x, u - x)dx & \text{si } 0 < u < 1 \\ \int_{u-1}^1 f_{X,Y}(x, u - x)dx & \text{si } 1 \leq u < 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} u & \text{si } 0 < u < 1 \\ 2 - u & \text{si } 1 \leq u < 2 \\ 0 & \text{en otro caso} \end{cases}$$

b. Se tiene $f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, v + x)dx$.

Ahora bien, para que podamos sustituir $f_{X,Y}(x, v+x)$ por 1 dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, v+x)dx$, se requiere tener $0 < x < 1$ y $0 < v+x < 1$, es decir $0 < x < 1$, $-v < x < 1-v$. Esta región se representa en la siguiente figura:

En la figura puede verse que, con el objeto de especificar el rango de valores de x dentro de la integral $\int_{-\infty}^{\infty} f_{X,Y}(x, v+x)dx$, conviene partir el rango de valores de u en dos intervalos, a saber $(-1, 0)$ y $[0, 1)$, obteniéndose:

$$f_V(v) = \begin{cases} \int_{-v}^1 f_{X,Y}(x, v+x)dx & \text{si } -1 < v < 0 \\ \int_0^{1-v} f_{X,Y}(x, v+x)dx & \text{si } 0 \leq v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} 1+v & \text{si } -1 < v < 0 \\ 1-v & \text{si } 0 \leq v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

EJEMPLO 2.9. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α_1, λ y α_2, λ respectivamente. Encuentre una función de densidad de $U = X + Y$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} y^{\alpha_2-1} e^{-\lambda(x+y)} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u-x)dx$$

$$\begin{aligned}
 &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda u} \int_0^u x^{\alpha_1-1} (u-x)^{\alpha_1-1} dx & \text{si } u > 0 \\ 0 & \text{en otro caso} \end{cases} \\
 &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda u} u^{\alpha_1+\alpha_2-1} \int_0^1 z^{\alpha_1-1} (1-z)^{\alpha_2-1} dz & \text{si } u > 0 \\ 0 & \text{en otro caso} \end{cases} \\
 &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} c e^{-\lambda u} u^{\alpha_1+\alpha_2-1} & \text{si } u > 0 \\ 0 & \text{en otro caso} \end{cases}
 \end{aligned}$$

en donde $c = \int_0^1 z^{\alpha_1-1} (1-z)^{\alpha_2-1} dz$.

Se puede concluir entonces que Z tiene distribución gama con parámetros $\alpha_1 + \alpha_2$ y λ .

COROLARIO 2.10. Sean X_1, X_2, \dots, X_n n variables aleatorias independientes, todas con distribución normal estándar, entonces la variable aleatoria $Z = X_1^2 + X_2^2 + \dots + X_n^2$ tiene distribución gama con parámetros $\alpha = \frac{n}{2}$ y $\lambda = \frac{1}{2}$. En particular, $X_1^2 + X_2^2$ tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$.

Demostración

Sabemos que, para cada $i \in \{1, \dots, n\}$, X_i^2 tiene distribución gama con parámetros $\frac{1}{2}$ y $\frac{1}{2}$, de manera que el resultado se sigue del último ejemplo. ■

COROLARIO 2.11. $\int_0^1 u^{\alpha_1-1} (1-u)^{\alpha_2-1} du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$

Demostración

De acuerdo con el ejemplo anterior, se tiene $\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} c = \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}$, de lo cual se obtiene el resultado. ■

DEFINICIÓN 2.12 (Función beta). La función $\beta : (0, \infty) \times (0, \infty) \mapsto \mathbb{R}$ definida por:

$$\beta(\alpha_1, \alpha_2) = \int_0^1 u^{\alpha_1-1} (1-u)^{\alpha_2-1} du$$

es llamada la función beta.

De acuerdo con el corolario 2.11, se tiene $\beta(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$.

EJEMPLO 2.13. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetros λ_1 y λ_2 respectivamente. Encuentre una función de densidad de $V = Y - X$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, v+x) dx = \begin{cases} \lambda_1 \lambda_2 e^{-\lambda_2 v} \int_{-\infty}^{\infty} e^{-(\lambda_1+\lambda_2)x} dx & \text{si } v < 0 \\ \lambda_1 \lambda_2 e^{-\lambda_2 v} \int_0^{\infty} e^{-(\lambda_1+\lambda_2)x} dx & \text{si } v \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} e^{\lambda_1 v} & \text{si } v < 0 \\ \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} e^{-\lambda_2 v} & \text{si } v \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

EJEMPLO 2.14. Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad $f_{X,Y}$. Encuentre una fórmula para una función de densidad de la variable aleatoria $Z = \frac{Y}{X} I_{[X \neq 0]}$.

Solución

Obsérvese que, como X es continua, $P[X \neq 0] = 1$.

$$\begin{aligned} F_Z(z) &= P\left[\frac{Y}{X} I_{[X \neq 0]} \leq z\right] = \iint_{\{(x,y) \in \mathbb{R}^2: \frac{y}{x} \leq z, x \neq 0\}} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^0 \int_{zx}^0 f_{X,Y}(x, y) dy dx + \int_0^{\infty} \int_0^{zx} f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^0 \int_z^0 x f_{X,Y}(x, ux) du dx + \int_0^{\infty} \int_0^z x f_{X,Y}(x, ux) du dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z |x| f_{X,Y}(x, ux) du dx \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} |x| f_{X,Y}(x, ux) dx du \end{aligned}$$

Por lo tanto:

$$\int_{-\infty}^{\infty} |x| f_{X,Y}(x, zx) dx$$

es una función de densidad de Z .

EJEMPLO 2.15. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetros λ_1 y λ_2 respectivamente. Encuentre una función de densidad de $Z = \frac{Y}{X}$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f_{X,Y}(x, zx) dx \\ &= \begin{cases} \lambda_1 \lambda_2 \int_0^{\infty} x e^{-(\lambda_1 + z\lambda_2)x} dx & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{\lambda_1 \lambda_2}{(\lambda_1 + z\lambda_2)^2} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

EJEMPLO 2.16. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentre una función de densidad de $Z = \frac{Y}{X} I_{[X \neq 0]}$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)},$$

para cualquier pareja $(x, y) \in \mathbb{R}^2$.

Se tiene entonces, para cualquier $z \in \mathbb{R}$:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f_{X,Y}(x, zx) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}(1+z^2)x^2} dx \\ &= \frac{1}{\pi} \int_0^{\infty} x e^{-\frac{1}{2}(1+z^2)x^2} dx = \frac{1}{\pi(1+z^2)} \int_0^{\infty} y e^{-\frac{1}{2}y^2} dy = \frac{1}{\pi(1+z^2)} \end{aligned}$$

Así que Z tiene distribución Cauchy.

EJEMPLO 2.17. Sean X y Y dos variables aleatorias independientes, X con distribución χ^2 con k grados de libertad, Y con distribución normal estándar. Encuentre una función de densidad de $Z = \frac{Y}{\sqrt{X/k}}$.

Solución

Se tiene:

$$f_{\sqrt{X}}(x, y) = \begin{cases} 2x f_X(x^2) & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{1}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})} x^{k-1} e^{-\frac{1}{2}x^2} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

De manera que una función de densidad conjunta de \sqrt{X} y Y está dada por:

$$f_{\sqrt{X}, Y}(x, y) = \begin{cases} \frac{1}{2^{\frac{k-1}{2}} \sqrt{\pi} \Gamma(\frac{k}{2})} x^{k-1} e^{-\frac{1}{2}(x^2+y^2)} & \text{si } x > 0, y \in \mathbb{R} \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$\begin{aligned} f_{\frac{Y}{\sqrt{X}}}(z) &= \int_{-\infty}^{\infty} |x| f_{\sqrt{X}, Y}(x, zx) dx = \frac{1}{2^{\frac{k-1}{2}} \sqrt{\pi} \Gamma(\frac{k}{2})} \int_0^{\infty} x^k e^{-\frac{1}{2}(1+z^2)x^2} dx \\ &= \frac{1}{2^{\frac{k-1}{2}} \sqrt{\pi} \Gamma(\frac{k}{2})} \int_0^{\infty} x^k e^{-\frac{1}{2}(1+z^2)x^2} dy = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi} \Gamma(\frac{k}{2})} \frac{1}{(1+z^2)^{\frac{k+1}{2}}} \end{aligned}$$

Por lo tanto:

$$f_Z(z) = \frac{1}{\sqrt{k}} f_{\frac{Y}{\sqrt{X}}}\left(\frac{z}{\sqrt{k}}\right) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \frac{1}{\left(1+\frac{z^2}{k}\right)^{\frac{k+1}{2}}}$$

DEFINICIÓN 2.18 (Distribución t). Se dice que la variable aleatoria X tiene distribución t con k grados de libertad si la función:

$$f_X(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \frac{1}{\left(1+\frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

es una función de densidad de X .

En la sección 3.4 demostraremos que si X_1, \dots, X_n son n variables aleatorias independientes, todas con distribución normal de parámetros μ y σ^2 , entonces la variable aleatoria $V = \frac{\sqrt{n}(\bar{X}-\mu)}{s_X}$ tiene distribución t con $n-1$ grados de libertad, en donde $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ y $s_X^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$. Por esta razón, una distribución t se utiliza cuando se quieren realizar estimaciones de la esperanza μ de una variable aleatoria con distribución normal de varianza desconocida.

A continuación se presentan las gráficas de algunas funciones de densidad t .

$k = 1$

$k = 5$

$k = 10$

Se puede observar en las gráficas que, a medida que k crece, una distribución t se va pareciendo más a una distribución normal estándar. Para ver esto más claramente, a continuación se muestran las gráficas de algunas funciones de densidad t , comparándolas con la función de densidad normal estándar. La línea punteada corresponde a la correspondiente densidad t , mientras que la línea sólida corresponde a la densidad normal estándar.

$k = 5$

$k = 10$

$k = 15$

$k = 20$

EJEMPLO 2.19. Sean X y Y dos variables aleatorias independientes, ambas con distribución χ^2 con m y n grados de libertad, respectivamente. Encuentre una función de densidad de $Z = \frac{Y/n}{X/m}$.

Solución

En este caso, una función de densidad conjunta de X y Y está dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{n}{2}) \Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{1}{2}(x+y)} & \text{si } x > 0, y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$\begin{aligned} f_{\frac{Y}{X}}(z) &= \int_{-\infty}^{\infty} |x| f_{X,Y}(x, zx) dx \\ &= \begin{cases} \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{n}{2}) \Gamma(\frac{m}{2})} z^{\frac{n}{2}-1} \int_0^{\infty} x^{\frac{m+n}{2}-1} e^{-\frac{1}{2}(1+z)x} dx & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \frac{z^{\frac{n}{2}-1}}{(1+z)^{\frac{m+n}{2}}} \int_0^{\infty} y^{\frac{m+n}{2}-1} e^{-y} dy & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \frac{z^{\frac{n}{2}-1}}{(1+z)^{\frac{m+n}{2}}} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

Por lo tanto:

$$\begin{aligned} f_Z(z) &= \frac{n}{m} f_{\frac{Y}{X}}\left(\frac{n}{m}z\right) = \begin{cases} \frac{n \Gamma(\frac{m+n}{2})}{m \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \frac{\left(\frac{nz}{m}\right)^{\frac{n}{2}-1}}{\left(1+\frac{nz}{m}\right)^{\frac{m+n}{2}}} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{\beta(\frac{m}{2}, \frac{n}{2})} \left(\frac{n}{m}\right)^{\frac{n}{2}} z^{\frac{n}{2}-1} \left(1 + \frac{n}{m}z\right)^{-\frac{m+n}{2}} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

DEFINICIÓN 2.20 (Distribución F). Se dice que la variable aleatoria X tiene distribución F con n y m grados de libertad si la función:

$$f_X(x) = \begin{cases} \frac{1}{\beta(\frac{m}{2}, \frac{n}{2})} \left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \left(1 + \frac{n}{m}x\right)^{-\frac{m+n}{2}} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X .

Basándonos en el ejemplo 2.19 y considerando que una variable aleatoria con distribución χ^2 se obtiene al sumar cuadrados de variables aleatorias con distribución normal estándar, una distribución F es básicamente la distribución del cociente de dos varianzas. Por esta razón, una distribución F se utiliza cuando se quiere comparar

las varianzas de dos poblaciones. Se utiliza también de esta manera en problemas relativos al **análisis de varianza**, en los cuales se trata de comparar las medias de 3 o más poblaciones. Para esto, se toma una muestra de cada población y, asumiendo que las muestras de cada población provienen todas ellas de una misma población, que incluye a las que están bajo estudio, se estima la varianza de esta población mediante dos procedimientos. Finalmente, las dos varianzas se comparan utilizando una distribución F.

A continuación se presentan las gráficas de algunas funciones de densidad F .

$$m = 1, n = 1$$

$$m = 3, n = 5$$

$$m = 10, n = 7$$

2.3. Distribuciones conjuntas de funciones de vectores aleatorios

La relación 1.2, utilizada en la sección anterior para encontrar una función de densidad de una función de un vector aleatorio absolutamente continuo, permite encontrar también la función de densidad conjunta de variables aleatorias que son funciones de un vector aleatorio absolutamente continuo.

EJEMPLO 2.21. Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Encuentre una fórmula para una función de densidad conjunta de las variables aleatorias $U = X + Y$ y $V = Y - X$.

Solución

$$\begin{aligned} P[U \leq u_0, V \leq v_0] &= P[X + Y \leq u_0, Y - X \leq v_0] \\ &= \iint_{\{(x,y):x+y \leq u_0, y-x \leq v_0\}} f_{X,Y}(x, y) dx dy \end{aligned}$$

Haciendo el cambio de variable $u = x + y$, $v = y - x$ se tiene $x = \frac{u-v}{2}$, $y = \frac{u+v}{2}$. Así que:

$$P[U \leq u_0, V \leq v_0] = \int_{-\infty}^{u_0} \int_{-\infty}^{u_0} \frac{1}{2} f_{X,Y} \left(\frac{u-v}{2}, \frac{u+v}{2} \right) dv du$$

Por lo tanto:

$$f_{X+Y, Y-X}(u, v) = \frac{1}{2} f_{X, Y} \left(\frac{u-v}{2}, \frac{u+v}{2} \right)$$

es una función de densidad conjunta de $X + Y$ y $Y - X$.

EJEMPLO 2.22. Sean X y Y variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad conjunta de $X + Y$ y $Y - X$ y utilícela para calcular $P \left[X + Y \leq \frac{3}{2}, Y - X \geq -\frac{1}{2} \right]$.

Solución

$$f_{X+Y, Y-X}(u, v) = \frac{1}{2} f \left(\frac{u-v}{2}, \frac{u+v}{2} \right) = \begin{cases} \frac{1}{2} & \text{si } 0 < \frac{u-v}{2} < 1, 0 < \frac{u+v}{2} < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{1}{2} & \text{si } 0 < u < 1, -u < v < u \text{ ó } 1 < u < 2, u-2 < v < -u+2 \\ 0 & \text{en otro caso} \end{cases}$$

El conjunto de puntos (u, v) para los cuales $f_{X+Y, Y-X}(u, v) > 0$ se representa en la siguiente figura:

Utilizando esta figura, se tiene:

$$P \left[X + Y \leq \frac{3}{2}, Y - X \geq -\frac{1}{2} \right] = P \left[U \leq \frac{3}{2}, V \geq -\frac{1}{2} \right]$$

$$= 1 - P \left[U > \frac{3}{2} \right] - P \left[V < -\frac{1}{2} \right] = 1 - \frac{1}{2} \int_{\frac{3}{2}}^2 \int_{u-2}^{-u+2} dv du - \frac{1}{2} \int_{-1}^{-\frac{1}{2}} \int_{-v}^{v+2} dudv$$

$$= 1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4}$$

▲

El método utilizado en el ejemplo 2.21 está basado en el teorema de cambio de variable para integrales múltiples. Este método puede utilizarse siempre que se cumplan las

condiciones para aplicar dicho teorema. De manera específica, se puede utilizar el siguiente resultado general.

PROPOSICIÓN 2.23. *Sea (X_1, \dots, X_n) un vector aleatorio absolutamente continuo con función de densidad conjunta f_{X_1, \dots, X_n} , $D \subset \mathbb{R}^n$ un conjunto abierto tal que $P[(X_1, \dots, X_n) \in D] = 1$ y $\varphi : \mathbb{R}^n \mapsto \mathbb{R}^n$ una función tal que:*

- (i) φ es inyectiva sobre D .
- (ii) Si $\phi : \varphi(D) \mapsto \mathbb{R}^n$ es la inversa de φ y ϕ_1, \dots, ϕ_n son las componentes de ϕ , entonces las derivadas parciales $\frac{\partial \phi_i}{\partial y_j}$ existen y son continuas sobre $\varphi(D)$.
- (iii) Si $J_\phi : \varphi(D) \mapsto \mathbb{R}$ es el Jacobiano de ϕ , entonces $J_\phi(y) \neq 0$ para cualquier $y \in \varphi(D)$.

Entonces el vector aleatorio $(Y_1, \dots, Y_n) = \varphi(X_1, \dots, X_n)$ es absolutamente continuo y una función de densidad conjunta, f_{Y_1, \dots, Y_n} , de (Y_1, \dots, Y_n) está dada por:

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = |J_\phi(y_1, \dots, y_n)| f_{X_1, \dots, X_n}(\phi(y_1, \dots, y_n))$$

para cualquier vector $(y_1, \dots, y_n) \in \mathbb{R}^n$.

EJEMPLO 2.24. *Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α_1, λ y α_2, λ , respectivamente. Encuentre una función de densidad de $\frac{X}{X+Y}$.*

Solución

La transformación $u = x, v = x + y$ tiene como inversa $x = u, y = v - u$, cuyo Jacobiano está dado por:

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

Por lo tanto:

$$f_{X, X+Y}(u, v) = f_{X, Y}(u, v-u) = \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1-1} (v-u)^{\alpha_2-1} e^{-\lambda v} & \text{si } u > 0, v-u > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u^{\alpha_1-1} (v-u)^{\alpha_2-1} e^{-\lambda v} & \text{si } 0 < u < v \\ 0 & \text{en otro caso} \end{cases}$$

Ahora bien, de acuerdo con el ejemplo 2.14, se tiene:

$$f_{\frac{X}{X+Y}}(z) = \int_{-\infty}^{\infty} |v| f_{X, X+Y}(vz, v) dv$$

$$\begin{aligned}
&= \begin{cases} \int_0^\infty v \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (vz)^{\alpha_1-1} (v-vz)^{\alpha_2-1} e^{-\lambda v} dv & \text{si } 0 < z < 1 \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} z^{\alpha_1-1} (1-z)^{\alpha_2-1} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty v^{\alpha_1+\alpha_2-1} e^{-\lambda v} dv & \text{si } 0 < z < 1 \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} z^{\alpha_1-1} (1-z)^{\alpha_2-1} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} & \text{si } 0 < z < 1 \\ 0 & \text{en otro caso} \end{cases} \\
&= \begin{cases} \frac{1}{\beta(\alpha_1, \alpha_2)} z^{\alpha_1-1} (1-z)^{\alpha_2-1} & \text{si } 0 < z < 1 \\ 0 & \text{en otro caso} \end{cases}
\end{aligned}$$

DEFINICIÓN 2.25 (Distribución beta). Se dice que la variable aleatoria X tiene distribución beta con parámetros α_1 y α_2 si la función:

$$f_X(x) = \begin{cases} \frac{1}{\beta(\alpha_1, \alpha_2)} x^{\alpha_1-1} (1-x)^{\alpha_2-1} & \text{si } x \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

es una función de densidad de X .

A continuación se presentan las gráficas de algunas funciones de densidad beta.

$$\alpha_1 = 1, \alpha_2 = 5$$

$$\alpha_1 = 3, \alpha_2 = 7$$

$$\alpha_1 = 6, \alpha_2 = 6$$

EJEMPLO 2.26. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α_1, λ y α_2, λ respectivamente. Encuentre una función de densidad conjunta de $U = X + Y$ y $V = \frac{X}{X+Y}$ y muestre que U y V son independientes.

Solución

La transformación $u = x + y$, $v = \frac{x}{x+y}$ tiene como inversa $x = uv$, $y = u - uv$, cuyo Jacobiano está dado por:

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = u$$

Por lo tanto:

$$\begin{aligned} f_{U,V}(u,v) &= f_{X+Y, \frac{X}{X+Y}}(u,v) = u f_{X,Y}(uv, u-uv) \\ &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} u(uv)^{\alpha_1-1} (u-uv)^{\alpha_2-1} e^{-\lambda u} & \text{si } uv > 0, u-uv > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} u^{\alpha_1+\alpha_2-1} e^{-\lambda u} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} v^{\alpha_1-1} (1-v)^{\alpha_2-1} & \text{si } u > 0, 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} u^{\alpha_1+\alpha_2-1} e^{-\lambda u} \frac{1}{\beta(\alpha_1, \alpha_2)} v^{\alpha_1-1} (1-v)^{\alpha_2-1} & \text{si } u > 0, 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases} \\ &= f_U(u) f_V(v) \end{aligned}$$

Así que U y V son independientes.

EJEMPLO 2.27. Sean X y Y variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad conjunta de $U = X$ y $V = XY$, grafique la región $\{(u, v) : f_{U,V}(u, v) > 0\}$ y calcule $P[U > \frac{1}{2}, V > \frac{1}{4}]$.

Solución

La transformación $u = x, v = xy$ tiene como inversa $x = u, y = \frac{v}{u}$, cuyo Jacobiano está dado por:

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} 1 & 0 \\ -\frac{v}{u^2} & \frac{1}{u} \end{vmatrix} = \frac{1}{u}$$

Por lo tanto:

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{u} f_{X,Y}(u, \frac{v}{u}) = \begin{cases} \frac{1}{u} & \text{si } 0 < u < 1, 0 < \frac{v}{u} < 1 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{u} & \text{si } 0 < v < u < 1 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

La región $\{(u, v) : f_{U,V}(u, v) > 0\}$ se representa en la siguiente figura:

Se tiene entonces:

$$P \left[U > \frac{1}{2}, V > \frac{1}{4} \right] = \int_{1/2}^1 \int_{\frac{1}{4}}^u \frac{1}{u} dv du = \frac{1}{2} - \frac{1}{4} \ln 2 = 0.32671$$

EJEMPLO 2.28. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Consideremos a la pareja (X, Y) como las coordenadas de un punto en el plano cartesiano y definamos R y θ como las coordenadas polares de ese punto. Encuentre una función de densidad conjunta de R y θ . ¿Son R y θ independientes?

Solución

El Jacobiano de la transformación $x = r \cos \theta$, $y = r \sin \theta$ está dado por r , de manera que se tiene:

$$f_{R,\theta}(r, \theta) = r f_{X,Y}(r \cos \theta, r \sin \theta) = \begin{cases} \frac{1}{2\pi} r e^{-\frac{1}{2}r^2} & \text{si } 0 < \theta < 2\pi, r > 0 \\ 0 & \text{en otro caso} \end{cases}$$

De aquí se sigue que:

$$f_R(r) = \int_0^{2\pi} f_{R,\theta}(r, \theta) d\theta = \begin{cases} r e^{-\frac{1}{2}r^2} & \text{si } r > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_\theta(\theta) = \int_0^\infty f_{R,\theta}(r, \theta) dr = \begin{cases} \frac{1}{2\pi} & \text{si } 0 < \theta < 2\pi \\ 0 & \text{en otro caso} \end{cases}$$

Así que $f_{R,\theta}(r, \theta) = f_R(r) f_\theta(\theta)$, por lo tanto, R y θ son independientes.

▲

Sean X, Y, R y θ como en el último ejemplo, entonces $Z = R^2 = X^2 + Y^2$ tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$ y, como R y θ son independientes, R^2 y θ también lo son. Además, se tiene $X = \sqrt{Z} \cos \theta$ y $Y = \sqrt{Z} \sin \theta$. Por otra parte, si U es una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$, entonces la variable aleatoria $-2 \ln U$ tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$. Esto sugiere el siguiente resultado:

Sean U y V dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$, entonces las variables aleatorias $X = \sqrt{-2 \ln U} \cos 2\pi V$ y $Y = \sqrt{-2 \ln U} \sin 2\pi V$ son independientes y ambas tienen distribución normal estándar.

La demostración de la validez de este resultado se deja como ejercicio.

Recuérdese que se dice que una familia $\{P_t : t \geq 0\}$ de variables aleatorias discretas forma un proceso de Poisson de parámetro λ si se satisfacen las siguientes propiedades:

- (i) $P_0 = 0$.
- (ii) Si $0 < t_1 < \dots < t_n$, entonces las variables aleatorias $P_{t_1}, P_{t_2} - P_{t_1}, \dots, P_{t_n} - P_{t_{n-1}}$ son independientes.
- (iii) Si $s < t$, entonces la variable aleatoria $P_t - P_s$ tiene distribución Poisson con parámetro $\lambda(t - s)$.

EJEMPLO 2.29. *Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t , entonces la familia de variables aleatorias $\{X_t\}_{t \geq 0}$ forma un proceso de Poisson de parámetro λ . Vamos a encontrar, para cada $n \in \mathbb{N}$, una función de densidad conjunta de T_1, \dots, T_n :*

Obsérvese primero que:

$$\begin{aligned} P[X_t = k] &= \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \frac{t^k}{k} = \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \int_0^t (t-s)^{k-1} ds \\ &= \int_0^t \frac{\lambda^k}{(k-1)!} (t-s)^{k-1} e^{-\lambda t} ds = \int_0^t \frac{[\lambda(t-s)]^{k-1}}{(k-1)!} e^{-\lambda(t-s)} \lambda e^{-\lambda s} ds \\ &= \int_0^t P[X_t - X_s = k-1] \lambda e^{-\lambda s} ds = \int_0^t P[X_{t-s} = k-1] \lambda e^{-\lambda s} ds \end{aligned}$$

Así que:

$$\begin{aligned} &P[X_{t_1} = k_1, \dots, X_{t_n} = k_n] \\ &= P[X_{t_1} = k_1, X_{t_2} - X_{t_1} = k_2 - k_1, \dots, X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}] \\ &= P[X_{t_1} = k_1] P[X_{t_2} - X_{t_1} = k_2 - k_1] \cdots P[X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}] \end{aligned}$$

$$= \int_0^{t_1} P[X_{t_1-s} = k_1 - 1] P[X_{t_2} - X_{t_1} = k_2 - k_1] \cdots$$

$$P[X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} P[X_{t_1-s} = k_1 - 1] P[X_{t_2-s} - X_{t_1-s} = k_2 - k_1] \cdots$$

$$P[X_{t_{n-s}} - X_{t_{n-1-s}} = k_n - k_{n-1}] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} P[X_{t_1-s} = k_1 - 1, \dots, X_{t_{n-s}} = k_n - 1] \lambda e^{-\lambda s} ds$$

Sean ahora $0 < t_1 < \dots < t_n$, entonces:

$$F_{T_1, \dots, T_n}(t_1, \dots, t_n) = P[T_1 \leq t_1, \dots, T_n \leq t_n] = P[X_{t_1} \geq 1, \dots, X_{t_n} \geq n]$$

$$= \sum_{\{k_1 \leq \dots \leq k_n : k_1 \geq 1, \dots, k_n \geq n\}} P[X_{t_1} = k_1, \dots, X_{t_n} = k_n]$$

$$= \int_0^{t_1} \sum_{\{k_1 \leq \dots \leq k_n : k_1 \geq 1, \dots, k_n \geq n\}} P[X_{t_1-s} = k_1 - 1, \dots, X_{t_n-s} = k_n - 1] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} \sum_{\{k_1 \leq \dots \leq k_n : k_1 \geq 0, \dots, k_n \geq n-1\}} P[X_{t_1-s} = k_1, \dots, X_{t_n-s} = k_n] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} \sum_{\{k_2 \leq \dots \leq k_n : k_2 \geq 1, \dots, k_n \geq n-1\}} P[X_{t_2-s} = k_2, \dots, X_{t_n-s} = k_n] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} P[X_{t_2-s} \geq 1, \dots, X_{t_n-s} \geq n-1] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} P[T_1 \leq t_2 - s, \dots, T_{n-1} \leq t_n - s] \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} F_{T_1, \dots, T_{n-1}}(t_2 - s, \dots, t_n - s) \lambda e^{-\lambda s} ds$$

Supongamos que $F_{T_1, \dots, T_{n-1}}$ admite una función de densidad $f_{T_1, \dots, T_{n-1}}$, entonces:

$$F_{T_1, \dots, T_{n-1}}(t_2 - s, \dots, t_n - s)$$

$$= \int \cdots \int_{\{0 < x_1 < \dots < x_{n-1} : x_1 \leq t_2 - s, \dots, x_{n-1} \leq t_n - s\}} f_{T_1, T_2, \dots, T_{n-1}}(x_1, \dots, x_{n-1}) dx_1 \cdots dx_{n-1}$$

$$= \int \cdots \int_{\{s < x_1 + s < \dots < x_{n-1} + s : x_1 + s \leq t_2, \dots, x_{n-1} + s \leq t_n\}} f_{T_1, T_2, \dots, T_{n-1}}(x_1, \dots, x_{n-1}) dx_1 \cdots dx_{n-1}$$

$$= \int \cdots \int_{\{s < y_2 < \dots < y_n : y_2 \leq t_2, \dots, y_n \leq t_n\}} f_{T_1, T_2, \dots, T_{n-1}}(y_2 - s, \dots, y_n - s) dy_2 \cdots dy_n$$

Así que:

$$F_{T_1, \dots, T_n}(t_1, \dots, t_n) = \int_0^{t_1} F_{T_1, \dots, T_{n-1}}(t_2 - s, \dots, t_n - s) \lambda e^{-\lambda s} ds$$

$$= \int_0^{t_1} \int \cdots \int_{\{s < y_2 < \dots < y_n : y_2 \leq t_2, \dots, y_n \leq t_n\}} f_{T_1, \dots, T_{n-1}}(y_2 - s, \dots, y_n - s) \lambda e^{-\lambda s} ds dy_2 \cdots dy_n$$

$$= \int \cdots \int_{\{0 < s < y_2 < \cdots < y_n : s \leq t_1, y_2 \leq t_2, \dots, y_n \leq t_n\}} f_{T_1, \dots, T_{n-1}}(y_2 - s, \dots, y_n - s) \lambda e^{-\lambda s} ds dy_2 \cdots dy_n$$

Por lo tanto, F_{T_1, T_2, \dots, T_n} admite como función de densidad a la función:

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = f_{T_1, \dots, T_{n-1}}(t_2 - t_1, \dots, t_n - t_1) \lambda e^{-\lambda t_1}$$

Por otra parte, se tiene:

$$P[T_1 \leq t_1] = P[X_{t_1} \geq 1] = 1 - e^{-\lambda t_1}$$

Así que F_{T_1} admite como función de densidad a la función:

$$f_{T_1}(t_1) = \begin{cases} \lambda e^{-\lambda t_1} & \text{si } t_1 > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$f_{T_1, T_2}(t_1, t_2) = f_{T_1}(t_2 - t_1) \lambda e^{-\lambda t_1} = \begin{cases} \lambda e^{-\lambda t_2} & \text{si } 0 < t_1 < t_2 \\ 0 & \text{en otro caso} \end{cases}$$

Un razonamiento de inducción muestra entonces que:

$$f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) = \begin{cases} \lambda^n e^{-\lambda t_n} & \text{si } 0 < t_1 < t_2 < \cdots < t_n \\ 0 & \text{en otro caso} \end{cases}$$

COROLARIO 2.30. Las variables aleatorias $Y_1 = T_1$, $Y_2 = T_2 - T_1$, $Y_3 = T_3 - T_2$, ... son independientes y todas tienen distribución exponencial con parámetro λ .

Demostración

$$\begin{aligned} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) &= f_{T_1, \dots, T_n}(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_n) \\ &= \begin{cases} \lambda^n e^{-\lambda(y_1 + \cdots + y_n)} & \text{si } y_1 > 0, y_2 > 0, \dots, y_n > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= f_{Y_1}(y_1) \cdots f_{Y_n}(y_n) \end{aligned}$$

■

LEMA 2.31. Para $0 \leq a < b$, se tiene $\int \cdots \int_{\{a < x_1 < \cdots < x_n \leq b\}} dx_1 \cdots dx_n = \frac{1}{n!} (b - a)^n$.

Demostración

$$\begin{aligned} \int \cdots \int_{\{s < x_1 < \cdots < x_n \leq t\}} dx_1 \cdots dx_n &= \int_s^t \cdots \int_{x_{n-2}}^t \int_{x_{n-1}}^t dx_n \cdots dx_1 \\ &= \int_s^t \cdots \int_{x_{n-2}}^t (t - x_{n-1}) dx_{n-1} \cdots dx_1 = \int_s^t \cdots \int_{x_{n-2}}^t \frac{1}{2} (t - x_{n-2})^2 dx_{n-2} \cdots dx_1 \\ &= \cdots = \int_s^t \frac{1}{(n-1)!} (t - x_2)^{n-1} dx_2 = \frac{1}{n!} (t - s)^n \end{aligned}$$

■

PROPOSICIÓN 2.32. *Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que las variables aleatorias $Y_1 = T_1$, $Y_2 = T_2 - T_1$, $Y_3 = T_3 - T_2, \dots$ son independientes y todas tienen distribución exponencial con parámetro λ . Para cada $t \geq 0$, sea X_t el número de veces que ocurre el evento hasta el tiempo t . Entonces, la familia de variables aleatorias $\{X_t\}_{t \geq 0}$ forma un proceso de Poisson de parámetro λ .*

Demostración

Para cualquier $n \in \mathbb{N}$ se tiene:

$$\begin{aligned} f_{T_1, \dots, T_n}(t_1, \dots, t_n) &= f_{Y_1, \dots, Y_n}(t_1, t_2 - t_1, \dots, t_n - t_{n-1}) \\ &= \begin{cases} \lambda^n e^{-\lambda t_n} & \text{si } 0 < t_1 < \dots < t_n \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

$$\text{Definamos } I_n(a, b) = \int \cdots \int_{\{a < x_1 < \dots < x_n \leq b\}} dx_1 \cdots dx_n$$

Para $0 < s < t$ y $m, k \in \{0, 1, \dots\}$, se tiene:

$$\begin{aligned} P[X_s = m, X_t - X_s = k] &= P[X_s = m, X_t = m + k] \\ &= P[T_m \leq s, T_{m+1} > s, T_{m+k} \leq t, T_{m+k+1} > t] \\ &= P[0 < T_1 < \dots < T_m \leq s, s < T_{m+1} \cdots < T_{m+k} \leq t, T_{m+k+1} > t] \\ &= \int \cdots \int_{\{0 < t_1 < t_2 < \dots < t_m \leq s, s < t_{m+1} < \dots < t_{m+k} \leq t\}} \int_t^\infty \lambda^{m+k+1} e^{-\lambda t_{m+k+1}} dt_{m+k+1} \cdots dt_1 \\ &= \lambda^{m+k} e^{-\lambda t} I_m(0, s) I_k(s, t) = \lambda^{m+k} e^{-\lambda t} \frac{1}{m!} s^m \frac{1}{k!} (t - s)^k \\ &= \frac{1}{m!} (\lambda s)^m e^{-\lambda s} \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)} \end{aligned}$$

Por lo tanto:

$$\begin{aligned} P[X_s = m] &= \sum_{k=0}^{\infty} P[X_s = m, X_t - X_s = k] \\ &= \frac{1}{m!} (\lambda s)^m e^{-\lambda s} \sum_{k=0}^{\infty} \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)} = \frac{1}{m!} (\lambda s)^m e^{-\lambda s} \\ P[X_t - X_s = k] &= \sum_{m=0}^{\infty} P[X_s = m, X_t - X_s = k] \\ &= \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)} \sum_{m=0}^{\infty} \frac{1}{m!} (\lambda s)^m e^{-\lambda s} = \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)} \\ P[X_t - X_s = k] &= \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)} \end{aligned}$$

Así que $X_t - X_s$ tiene distribución Poisson con parámetro $\lambda(t - s)$.

De la misma forma, para $0 < t_1 < \dots < t_n$ y $k_1, \dots, k_n \in \{0, 1, \dots\}$, se tiene:

$$\begin{aligned}
& P [X_{t_1} = k_1, X_{t_2} - X_{t_1} = k_2, \dots, X_{t_n} - X_{t_{n-1}} = k_n] \\
& P [X_{t_1} = k_1, X_{t_2} = k_1 + k_2, \dots, X_{t_n} = k_1 + k_2 + \dots + k_n] \\
& = P [T_{k_1} \leq t_1 < T_{k_1+1}, T_{k_1+k_2} \leq t_2 < T_{k_1+k_2+1}, \dots, T_{k_1+\dots+k_n} \leq t_n < T_{k_1+\dots+k_n+1}] \\
& = P [T_{k_1} \leq t_1, t_1 < T_{k_1+1} < T_{k_1+k_2} \leq t_2, \dots, \\
& t_{n-1} < T_{k_1+\dots+k_{n-1}+1} < T_{k_1+\dots+k_n} \leq t_n, T_{k_1+\dots+k_n+1} > t_n] \\
& = \lambda^{k_1+\dots+k_n} e^{-\lambda t_n} I_{k_1}(0, t_1) I_{k_2}(t_1, t_2) \dots I_{k_n}(t_{n-1}, t_n) \\
& = \lambda^{k_1+\dots+k_n} e^{-\lambda t_n} \frac{1}{k_1!} t_1^{k_1} \frac{1}{k_2!} (t_2 - t_1)^{k_2} \dots \frac{1}{k_n!} (t_n - t_{n-1})^{k_n} \\
& = \frac{1}{k_1!} (\lambda t_1)^{k_1} e^{-\lambda t_1} \frac{1}{k_2!} [\lambda(t_2 - t_1)]^{k_2} e^{-\lambda(t_2-t_1)} \dots \frac{1}{k_n!} [\lambda(t_n - t_{n-1})]^{k_n} e^{-\lambda(t_n-t_{n-1})}
\end{aligned}$$

Así que:

$$\begin{aligned}
& P [X_{t_1} = k_1, X_{t_2} - X_{t_1} = k_2, \dots, X_{t_n} - X_{t_{n-1}} = k_n] \\
& = P [X_{t_1} = k_1] P [X_{t_2} - X_{t_1} = k_2] \dots P [X_{t_n} - X_{t_{n-1}} = k_n]
\end{aligned}$$

Por lo tanto, las variables aleatorias $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes.

2.4. Estadísticos de orden

DEFINICIÓN 2.33 (Estadísticos de orden). Sean X_1, \dots, X_n n variables aleatorias. Las variables aleatorias, $X_{(1)}, \dots, X_{(n)}$, las cuales se obtienen ordenando X_1, \dots, X_n en forma creciente, son llamadas los estadísticos de orden correspondientes a X_1, \dots, X_n .

PROPOSICIÓN 2.34. Sean X_1, \dots, X_n n variables aleatorias absolutamente continuas e independientes, con función de densidad común f . Entonces una función de densidad conjunta de los estadísticos de orden, $X_{(1)}, \dots, X_{(n)}$, correspondientes a X_1, \dots, X_n , está dada por:

$$f_{X_{(1)}, \dots, X_{(n)}}(u_1, \dots, u_n) = \begin{cases} n! f(u_1) \dots f(u_n) & \text{si } u_1 < \dots < u_n \\ 0 & \text{en otro caso} \end{cases}$$

Demostración

Sea $(u_1, \dots, u_n) \in \mathbb{R}^n$, entonces:

$$\begin{aligned}
F_{X_{(1)}, \dots, X_{(n)}}(u_1, \dots, u_n) &= P[X_{(1)} \leq u_1, \dots, X_{(n)} \leq u_n] \\
&= P[X_1 \leq u_1, \dots, X_n \leq u_n, X_1 < \dots < X_n] \\
&+ P[X_2 \leq u_1, \dots, X_n \leq u_n, X_2 < \dots < X_n] \\
&+ \dots \\
&= \int_{\{(x_1, \dots, x_n): x_1 \leq u_1, \dots, x_n \leq u_n, x_1 < \dots < x_n\}} f(x_1) \cdots f(x_n) dx_1 \cdots dx_n \\
&+ \int_{\{(x_1, \dots, x_n): x_2 \leq u_1, \dots, x_n \leq u_n, x_2 < \dots < x_n\}} f(x_2) f(x_1) \cdots f(x_n) dx_2 dx_1 \cdots dx_n \\
&+ \dots \\
&= n! \int_{\{(x_1, \dots, x_n): x_1 \leq u_1, \dots, x_n \leq u_n, x_1 < \dots < x_n\}} f(x_1) \cdots f(x_n) dx_1 \cdots dx_n \\
&= \int_{\{(x_1, \dots, x_n): x_1 \leq u_1, \dots, x_n \leq u_n\}} I_{\{(y_1, \dots, y_n) \in \mathbb{R}^n: y_1 < \dots < y_n\}}(x_1, \dots, x_n) n! f(x_1) \cdots f(x_n) dx_1 \cdots dx_n
\end{aligned}$$

Por lo tanto:

$$f_{X_{(1)}, \dots, X_{(n)}}(u_1, \dots, u_n) = \begin{cases} n! f(u_1) \cdots f(u_n) & \text{si } u_1 < \dots < u_n \\ 0 & \text{en otro caso} \end{cases}$$

■

EJEMPLO 2.35. Sean T_1, T_2 y T_3 tres variables aleatorias independientes, todas con distribución exponencial de parámetro λ . Encuentre la probabilidad de que ningún par de ellas difiera en menos de t , en donde $t > 0$.

Solución

Sean $T_{(1)}, T_{(2)}, T_{(3)}$ los estadísticos de orden correspondientes a T_1, T_2, T_3 , entonces:

$$f_{T_{(1)}, T_{(2)}, T_{(3)}}(t_1, t_2, t_3) = \begin{cases} 3! \lambda^3 e^{-\lambda(t_1+t_2+t_3)} & \text{si } 0 < t_1 < t_2 < t_3 \\ 0 & \text{en otro caso} \end{cases}$$

$$\begin{aligned}
&P[|T_2 - T_1| \geq t, |T_3 - T_1| \geq t, |T_3 - T_2| \geq t] \\
&= P[T_{(2)} \geq T_{(1)} + t, T_{(3)} \geq T_{(2)} + t] \\
&= \int_0^\infty \int_{t_1+t}^\infty \int_{t_2+t}^\infty f_{T_{(1)}, T_{(2)}, T_{(3)}}(t_1, t_2, t_3) dt_3 dt_2 dt_1 \\
&= \int_0^\infty \int_{t_1+t}^\infty \int_{t_2+t}^\infty 3! \lambda^3 e^{-\lambda(t_1+t_2+t_3)} dt_3 dt_2 dt_1
\end{aligned}$$

$$\begin{aligned}
&= 6 \int_0^\infty \int_{t_1+t}^\infty e^{-\lambda t} \lambda^2 e^{-\lambda(t_1+2t_2)} dt_2 dt_1 \\
&= 3 \int_0^\infty e^{-3\lambda t} \lambda e^{-3\lambda t_1} dt_1 = e^{-3\lambda t}
\end{aligned}$$

EJEMPLO 2.36. *Se seleccionan, al azar y de manera independiente, n puntos en el intervalo $(0, 1)$. Encuentre la probabilidad de que no haya dos de ellos cuya distancia sea menor que d , en donde $0 < d \leq \frac{1}{n-1}$.*

Solución

Sean X_1, \dots, X_n los n puntos seleccionados y $X_{(1)}, \dots, X_{(n)}$ los estadísticos de orden correspondientes a X_1, \dots, X_n , entonces:

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n! & \text{si } 0 < x_1 < \dots < x_n < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\begin{aligned}
&P[|X_i - X_j| \geq d \text{ para cualquier pareja } i, j \in \{1, \dots, n\} \text{ con } i \neq j] \\
&= P[X_{(2)} \geq X_{(1)} + d, \dots, X_{(n)} \geq X_{(n-1)} + d] \\
&= \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_{x_{n-2}+d}^{1-d} \int_{x_{n-1}+d}^1 f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) dx_n \dots dx_1 \\
&= \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_{x_{n-2}+d}^{1-d} \int_{x_{n-1}+d}^1 n! dx_n dx_{n-1} \dots dx_1 \\
&= n! \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_{x_{n-2}+d}^{1-d} (1 - x_{n-1} - d) dx_{n-1} \dots dx_1 \\
&= n! \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_0^{1-2d-x_{n-2}} y_{n-1} dy_{n-1} \dots dx_1 \\
&= n! \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_{x_{n-3}+d}^{1-2d} \frac{1}{2} (1 - 2d - x_{n-2})^2 dx_{n-2} \dots dx_1 \\
&= n! \int_0^{1-(n-1)d} \int_{x_1+d}^{1-(n-2)d} \dots \int_{x_{n-4}+d}^{1-3d} \frac{1}{3!} (1 - 3d - x_{n-3})^3 dx_{n-3} \dots dx_1 \\
&= \dots = n! \int_0^{1-(n-1)d} \frac{1}{(n-1)!} [1 - (n-1)d - x_1]^{n-1} dx_1 = [1 - (n-1)d]^n
\end{aligned}$$

PROPOSICIÓN 2.37. *Sean X_1, \dots, X_n n variables aleatorias absolutamente continuas e independientes, con función de densidad común f . Sea F la función de distribución común de X_1, \dots, X_n , $a = \inf \{x \in \mathbb{R} : F(x) > 0\}$ y $b = \sup \{x \in \mathbb{R} : F(x) < 1\}$. Supongamos que f es continua en el intervalo (a, b) de tal manera que cuando a (resp. b) es finito, f se puede extender continuamente a a (resp. b), entonces:*

a) Para $k \in \{1, \dots, n\}$, las funciones de distribución y de densidad del k -ésimo estadístico de orden, $X_{(k)}$, están dadas, respectivamente, por:

$$F_{X_{(k)}} = \sum_{j=k}^n \binom{n}{j} F^j [1 - F]^{n-j}$$

$$f_{X_{(k)}} = k \binom{n}{k} F^{k-1} [1 - F]^{n-k} f$$

b) Una función de densidad conjunta de $X_{(1)}$ y $X_{(n)}$ está dada por:

$$f_{X_{(1)}, X_{(n)}}(x, y) = \begin{cases} n(n-1) [F(y) - F(x)]^{n-2} f(x)f(y) & \text{si } x < y \\ 0 & \text{en otro caso} \end{cases}$$

Demostración

a. El evento $[X_{(k)} \leq x]$ ocurre cuando ocurren k o más de los eventos $[X_j \leq x]$, en donde $j \in \{1, \dots, n\}$. Por otra parte, el número de ocurrencias de los eventos $[X_j \leq x]$, en donde $j \in \{1, \dots, n\}$, tiene distribución binomial con parámetros n y $p = F(x)$. Por lo tanto:

$$F_{X_{(k)}}(x) = P[X_{(k)} \leq x] = \sum_{j=k}^n \binom{n}{j} F^j(x) [1 - F(x)]^{n-j}$$

Así que, para $x \in (a, b)$, se tiene:

$$\begin{aligned} F'_{X_{(k)}}(x) &= \sum_{j=k}^n \binom{n}{j} j F^{j-1}(x) [1 - F(x)]^{n-j} f(x) \\ &- \sum_{j=k}^{n-1} \binom{n}{j} (n-j) F^j(x) [1 - F(x)]^{n-j-1} f(x) \\ &= k \binom{n}{k} F^{k-1}(x) [1 - F(x)]^{n-k} f(x) \\ &+ \sum_{j=k+1}^n \frac{n!}{(j-1)!(n-j)!} F^{j-1}(x) [1 - F(x)]^{n-j} f(x) \\ &- \sum_{j=k}^{n-1} \frac{n!}{j!(n-j-1)!} F^j(x) [1 - F(x)]^{n-j-1} f(x) \\ &= k \binom{n}{k} F^{k-1}(x) [1 - F(x)]^{n-k} f(x) \\ &+ \sum_{j=k}^{n-1} \frac{n!}{j!(n-j-1)!} F^j(x) [1 - F(x)]^{n-j-1} f(x) - \sum_{j=k}^{n-1} \frac{n!}{j!(n-j-1)!} F^j(x) [1 - F(x)]^{n-j-1} f(x) \\ &= k \binom{n}{k} F^{k-1}(x) [1 - F(x)]^{n-k} f(x) \end{aligned}$$

Por lo tanto:

$$f_{X_{(k)}} = F'_{X_{(k)}} = k \binom{n}{k} F^{k-1} [1 - F]^{n-k} f.$$

b. Sea $x < y$, entonces:

$$P[X_{(1)} > x, X_{(n)} \leq y] = P[x < X_1 \leq y, x < X_2 \leq y, \dots, x < X_n \leq y]$$

$$= [F(y) - F(x)]^n$$

Así que:

$$\begin{aligned} F_{X_{(1)}, X_{(n)}}(x, y) &= P[X_{(1)} \leq x, X_{(n)} \leq y] = P[X_{(n)} \leq y] - P[X_{(1)} > x, X_{(n)} \leq y] \\ &= F^n(y) - [F(y) - F(x)]^n \end{aligned}$$

de lo cual se sigue el resultado. ■

COROLARIO 2.38. Sean X_1, \dots, X_n n variables aleatorias absolutamente continuas e independientes, con función de densidad común f . Sea F la función de distribución común de X_1, \dots, X_n , $a = \inf\{x \in \mathbb{R} : F(x) > 0\}$ y $b = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Supongamos que f es continua en el intervalo (a, b) , entonces:

$$F_{\min\{X_1, \dots, X_n\}} = 1 - [1 - F]^n$$

$$f_{\min\{X_1, \dots, X_n\}} = n[1 - F]^{n-1} f$$

$$F_{\max\{X_1, \dots, X_n\}} = F^n$$

$$f_{\max\{X_1, \dots, X_n\}} = nF^{n-1} f$$

EJEMPLO 2.39. Dada una cierta producción de lámparas, se sabe que el tiempo de vida, en horas, de cada una de ellas es independiente del tiempo de vida de las otras y tiene distribución exponencial con parámetro λ . Si se prenden n lámparas simultáneamente, ¿cuál es la probabilidad de que a) ninguna lámpara esté funcionando después de n horas?, b) la primera lámpara deje de funcionar dentro de la primera hora?, c) dejen de funcionar 3 o más lámparas dentro de la primera hora?, d) dejen de funcionar exactamente 3 lámparas dentro de la primera hora? e) Si T es el tiempo que transcurre desde que deja de funcionar la primera lámpara hasta que deja de funcionar la última, encuentre la distribución de T .

Solución

Para $i \in \{1, \dots, n\}$, sea T_i el tiempo de vida de la i -ésima lámpara y sean F y f la función de distribución y una función de densidad común, respectivamente, de T_1, \dots, T_n .

$$a. P[T_1 \leq n, \dots, T_n \leq n] = P[\max\{T_1, \dots, T_n\} \leq n]$$

$$= F^n(n) = (1 - e^{-\lambda n})^n$$

$$b. P[\min\{T_1, \dots, T_n\} < 1] = 1 - [1 - F(1)]^n = 1 - e^{-\lambda n}$$

$$\begin{aligned}
c. P [T_{(3)} < 1] &= \sum_{j=3}^n \binom{n}{j} F^j(1) [1 - F(1)]^{n-j} \\
&= 1 - \sum_{j=0}^2 \binom{n}{j} F^j(1) [1 - F(1)]^{n-j} = 1 - \sum_{j=0}^2 \binom{n}{j} (1 - e^{-\lambda})^j e^{-\lambda(n-j)} \\
&= 1 - e^{-\lambda n} - n(1 - e^{-\lambda})e^{-\lambda(n-1)} - \frac{1}{2}n(n-1)(1 - e^{-\lambda})^2 e^{-\lambda(n-2)} \\
&= 1 - \frac{1}{2}(n-1)(n-2)e^{-\lambda n} + n(n-2)e^{-\lambda(n-1)} - \frac{1}{2}n(n-1)e^{-\lambda(n-2)}
\end{aligned}$$

$$\begin{aligned}
d. P [T_{(3)} < 1, T_{(4)} \geq 1] &= P [T_{(3)} < 1] - P [T_{(4)} < 1] \\
&= \binom{n}{3} F^3(1) [1 - F(1)]^{n-3} = \binom{n}{3} (1 - e^{-\lambda})^3 e^{-\lambda(n-3)}
\end{aligned}$$

e. Para $t > 0$, se tiene:

$$\begin{aligned}
f_T(t) &= f_{T_{(n)} - T_{(1)}}(t) = \int_{-\infty}^{\infty} f_{T_{(1)}, T_{(n)}}(x, x+t) dx \\
&= \int_0^{\infty} n(n-1) [F(x+t) - F(x)]^{n-2} f(x) f(x+t) dx \\
&= \int_0^{\infty} n(n-1) \lambda^2 e^{-\lambda x(n-2)} (1 - e^{-\lambda t})^{n-2} e^{-2\lambda x} e^{-\lambda t} dx \\
&= (n-1) \lambda e^{-\lambda t} (1 - e^{-\lambda t})^{n-2} \int_0^{\infty} \lambda n e^{-\lambda n x} dx = (n-1) \lambda e^{-\lambda t} (1 - e^{-\lambda t})^{n-2}
\end{aligned}$$

2.5. Esperanza de funciones de vectores aleatorios

En esta sección vamos a generalizar al caso vectorial algunos resultados que fueron expuestos en el capítulo 9 del primer volumen de este libro.

Recordemos que se dice que una variable aleatoria discreta X tiene esperanza finita si la serie $\sum_x |x| f_X(x)$ converge, en cuyo caso se define la esperanza de X , $E[X]$, mediante la fórmula:

$$E[X] = \sum_x x f_X(x)$$

De la misma manera, se dice que una variable aleatoria absolutamente continua X tiene esperanza finita si la integral $\int_{-\infty}^{\infty} |x| f_X(x) dx$ es finita, en cuyo caso se define la esperanza de X , $E[X]$, mediante la fórmula:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

De manera general, si X es cualquier variable aleatoria con función de distribución F_X , se dice que X tiene esperanza finita si $\int_0^{\infty} P[X > x] dx < \infty$ y $\int_{-\infty}^0 P[X < x] dx < \infty$ y, en ese caso, se define la esperanza de X , $E[X]$, mediante la fórmula:

$$\begin{aligned} E[X] &= \int_0^\infty P[X > x] dx - \int_{-\infty}^0 P[X < x] dx \\ &= \int_0^\infty [1 - F_X(x)] dx - \int_{-\infty}^0 F_X(x) dx \end{aligned}$$

Finalmente, recordemos que se tienen los siguientes resultados, los cuales fueron demostrados en el primer volumen (corolarios 9.43 y 9.44):

PROPOSICIÓN 2.40. *Sean X_1, \dots, X_n n variables aleatorias de esperanza finita, entonces $\sum_{k=1}^n X_k$ también tiene esperanza finita y $E[\sum_{k=1}^n X_k] = \sum_{k=1}^n E[X_k]$.*

PROPOSICIÓN 2.41. *Sean X_1, \dots, X_n n variables aleatorias independientes de esperanza finita, entonces $\prod_{k=1}^n X_k$ también tiene esperanza finita y $E[\prod_{k=1}^n X_k] = \prod_{k=1}^n E[X_k]$.*

Los siguientes resultados nos van a permitir encontrar una fórmula simple para calcular la esperanza de una función de un vector aleatorio discreto o absolutamente continuo.

PROPOSICIÓN 2.42. *Sea X un vector aleatorio n -dimensional discreto con función de densidad conjunta f_X , $g: \mathbb{R}^n \mapsto \mathbb{R}$ cualquier función boreliana no negativa y $F_{g(X)}$ la función de distribución de $g(X)$. Entonces:*

$$\int_0^\infty P[g(X) > y] dz = \sum_{\{x \in V_X\}} g(x) f_X(x)$$

en donde V_X es el conjunto de posibles valores de X .

Demostración

$$\begin{aligned} \int_0^\infty P[g(X) > y] dy &= \int_0^\infty \sum_{\{x \in V_X: g(x) > y\}} f_X(x) dy \\ &= \int_0^\infty \sum_{\{x \in V_X\}} I_{[0, g(x))}(y) f_X(x) dy = \sum_{\{x \in V_X\}} \int_0^\infty I_{[0, g(x))}(y) f_X(x) dy \\ &= \sum_{\{x \in V_X\}} g(x) f_X(x) \end{aligned}$$

■

COROLARIO 2.43. *Sea X un vector aleatorio n -dimensional discreto con función de densidad conjunta f_X , $g: \mathbb{R}^n \mapsto \mathbb{R}$ cualquier función boreliana. Entonces, $g(X)$ tiene esperanza finita si y sólo si $\sum_{x \in V_X} |g(x)| f_X(x) < \infty$ y, en ese caso, se tiene:*

$$E[g(X)] = \sum_{x \in V_X} g(x) f_X(x)$$

en donde V_X es el conjunto de posibles valores de X .

PROPOSICIÓN 2.44. *Sea X un vector aleatorio n -dimensional absolutamente continuo, con función de densidad conjunta f_X y $g : \mathbb{R}^n \mapsto \mathbb{R}$ cualquier función boreliana no negativa. Entonces:*

$$\int_0^\infty P[g(X) > y] dy = \int \cdots \int_{\mathbb{R}^n} g(x) f_X(x) dx$$

Demostración

$$\begin{aligned} \text{a. } \int_0^\infty P[g(X) > y] dy &= \int_0^\infty \int \cdots \int_{\{x \in \mathbb{R}^n : g(x) > y\}} f_X(x) dx dy \\ &= \int_0^\infty \int \cdots \int_{\mathbb{R}^n} I_{[0, g(x))}(y) f_X(x) dx dy \\ &= \int \cdots \int_{\mathbb{R}^n} \int_0^\infty I_{[0, g(x))}(y) f_X(x) dy dx \\ &= \int \cdots \int_{\mathbb{R}^n} g(x) f_X(x) dx \end{aligned}$$

■

COROLARIO 2.45. *Sea X un vector aleatorio n -dimensional absolutamente continuo, con función de densidad conjunta f_X y $g : \mathbb{R}^n \mapsto \mathbb{R}$ cualquier función boreliana. Entonces, $g(X)$ tiene esperanza finita si y sólo si $\int \cdots \int_{\mathbb{R}^n} |g(x)| f_X(x) dx < \infty$, en ese caso, se tiene:*

$$E[g(X)] = \int \cdots \int_{\mathbb{R}^n} g(x) f_X(x) dx$$

EJEMPLO 2.46. *Sean X, Y y Z tres variables aleatorias independientes, todas con distribución exponencial de parámetro λ . Encuentre a) $E[\max\{X, Y, Z\}]$ y b) $E[\min\{X, Y, Z\}]$.*

Solución

$$\begin{aligned} \text{a. } E[\max\{X, Y, Z\}] &= \int_0^\infty \int_0^\infty \int_0^\infty \max\{x, y, z\} f_{X,Y,Z}(x, y, z) dx dy dz \\ &= 6 \iiint_{\{(x,y,z): x < y < z\}} \max\{x, y, z\} f_X(x) f_Y(y) f_Z(z) dx dy dz \\ &= 6 \iiint_{\{(x,y,z): x < y < z\}} z \lambda^3 e^{-\lambda(x+y+z)} dx dy dz = 6 \int_0^\infty \int_0^z \int_0^y \lambda^3 z e^{-\lambda(x+y+z)} dx dy dz = \frac{11}{6\lambda} \\ \text{b. } E[\min\{X, Y, Z\}] &= \int_0^\infty \int_0^\infty \int_0^\infty \min\{x, y, z\} f_{X,Y,Z}(x, y, z) dx dy dz \\ &= 6 \iiint_{\{(x,y,z): x < y < z\}} \min\{x, y, z\} f_X(x) f_Y(y) f_Z(z) dx dy dz \\ &= 6 \iiint_{\{(x,y,z): x < y < z\}} x \lambda^3 e^{-\lambda(x+y+z)} dx dy dz = 6 \int_0^\infty \int_x^\infty \int_y^\infty \lambda^3 x e^{-\lambda(x+y+z)} dz dy dx = \frac{1}{3\lambda} \end{aligned}$$

2.5.1. Coeficiente de correlación y matriz de covarianzas. Recordemos que si X es una variable aleatoria de esperanza finita, se define la varianza de X , $Var(X)$, mediante la relación:

$$Var(X) = E[(X - E(X))^2] = E[X^2] - (E[X])^2$$

También, si X y Y son dos variables aleatorias de varianza finita, entonces XY tiene esperanza finita y se define la covarianza de X y Y , $Cov(X, Y)$, mediante la relación:

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

Las siguientes dos proposiciones fueron demostradas en el primer volumen de este libro. Para comodidad en la lectura, se reproducen a continuación:

PROPOSICIÓN 2.47 (Desigualdad de Cauchy-Schwarz). Sean X y Y dos variables aleatorias cualesquiera, entonces:

$$E[|XY|] \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

Además, si X y Y tienen varianza finita, entonces $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$ si y sólo si existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$.

Demostración

Si $E[X^2] = \infty$ o $E[Y^2] = \infty$ la desigualdad es obvia.

Supongamos ahora que $E[X^2] < \infty$ y $E[Y^2] < \infty$, es decir, que tanto X como Y tienen varianza finita.

Sea $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$.

Si $\alpha = 0$, se tiene $E[Y^2] = 0$, de manera que:

$$P[|XY| = 0] \geq P[X = 0] = P[X^2 = 0] = 1$$

Por lo tanto, $E[|XY|] = 0$. Así que se cumple la desigualdad.

De la misma manera, si $\beta = 0$, entonces $E[|XY|] = 0$. Así que se cumple la desigualdad.

Supongamos ahora que $\alpha > 0$ y $\beta > 0$.

Sabemos que $\alpha|X| - \beta|Y|$ tiene varianza finita y se tiene:

$$0 \leq E[(\alpha|X| - \beta|Y|)^2] = \alpha^2 E[X^2] + \beta^2 E[Y^2] - 2\alpha\beta E[|XY|] = 2\alpha^2\beta^2 - 2\alpha\beta E[|XY|]$$

Así que, $\alpha\beta - E[|XY|] \geq 0$. Es decir, $E[|XY|] \leq \alpha\beta$.

Para la segunda parte, supongamos primero que X y Y tienen varianza finita y que $|E[XY]| = \sqrt{E[X^2]}\sqrt{E[Y^2]}$.

Definiendo, como antes, $\alpha = (E[Y^2])^{\frac{1}{2}}$ y $\beta = (E[X^2])^{\frac{1}{2}}$, se tiene:

Si $\alpha = 0$ y $\beta = 0$, entonces $P[X = 0] = P[Y = 0] = 1$. Por lo tanto $P[X = 0, Y = 0] = 1$. De manera que, tomando en consideración que $P[X = 0, Y = 0] \leq P[X + Y = 0]$, se tiene $P[X + Y = 0] = 1$. Es decir, se tiene el resultado deseado con $a = b = 1$.

Si $\alpha \neq 0$ ó $\beta \neq 0$ se tienen los siguientes dos casos:

Si $E[XY] > 0$, entonces:

$$0 \leq E[(\alpha X - \beta Y)^2] = 2\alpha^2\beta^2 - 2\alpha\beta E[XY] = 0$$

Así que, $E[(\alpha X - \beta Y)^2] = 0$, de lo cual se sigue $P[\alpha X - \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = -\beta$.

Si $E[XY] < 0$, entonces:

$$0 \leq E[(\alpha X + \beta Y)^2] = 2\alpha^2\beta^2 + 2\alpha\beta E[XY] = 0$$

Así que, $E[(\alpha X + \beta Y)^2] = 0$, de lo cual se sigue $P[\alpha X + \beta Y = 0] = 1$.

Es decir, se tiene el resultado deseado con $a = \alpha$ y $b = \beta$.

Finalmente, supongamos que existen constantes a y b tales que por lo menos una de ellas es distinta de cero y $P[aX + bY = 0] = 1$. Supongamos, por ejemplo, que $a \neq 0$, entonces $P[X = -\frac{b}{a}Y] = 1$. Así que:

$$(E[XY])^2 = \frac{b^2}{a^2} (E[Y^2])^2 = E\left[\left(-\frac{b}{a}Y\right)^2\right] E[Y^2] = E[X^2] E[Y^2]$$

■

PROPOSICIÓN 2.48. Sean X y Y dos variables aleatorias de varianza finita. Entonces:

$$|Cov(X, Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}$$

Además, la igualdad se cumple si y sólo si existen constantes a , b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$.

Demostración

Utilizando la proposición 2.47, se tiene:

$$\begin{aligned} |Cov(X, Y)| &= |E[(X - E[X])(Y - E[Y])]| \leq E[|X - E[X]| |Y - E[Y]|] \\ &\leq \sqrt{E[(X - E[X])^2]} \sqrt{E[(Y - E[Y])^2]} = \sqrt{Var(X)} \sqrt{Var(Y)} \end{aligned}$$

Si la igualdad se cumple, entonces se tiene:

$$|E[(X - E[X])(Y - E[Y])]| = \sqrt{E[(X - E[X])^2]} \sqrt{E[(Y - E[Y])^2]}$$

De manera que, nuevamente por la proposición 2.47, existen constantes a y b tales que no son ambas cero y $P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$. Es decir, $P[aX + bY = c] = 1$, en donde $c = aE[X] + bE[Y]$.

Supongamos ahora que existen constantes a , b y c tales que a y b no son ambas cero y $P[aX + bY = c] = 1$. Entonces $E[aX + bY - c] = 0$, de lo cual se sigue $c = E[aX + bY]$. De manera que se tiene:

$$P[a(X - E[X]) + b(Y - E[Y]) = 0] = 1$$

Así que, por la proposición 2.47, se tiene:

$$|Cov(X, Y)| = |E[(X - E[X])(Y - E[Y])]| = \sqrt{Var(X)} \sqrt{Var(Y)}$$

■

DEFINICIÓN 2.49 (Coeficiente de correlación). Sean X y Y dos variables aleatorias de varianza finita y positiva. Se define el coeficiente de correlación, $\rho_{X,Y}$, mediante la relación:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

De la proposición 2.48 se sigue inmediatamente que, para cualquier par de variables aleatorias de varianza finita y positiva, $-1 \leq \rho_{X,Y} \leq 1$.

En la demostración de la proposición 2.47, de la cual se sigue 2.48, se puede ver que si $|\rho_{X,Y}| = 1$ entonces existen constantes a , b y c tales que a y b son positivas y $P[aX - bY = c] = 1$ cuando $\rho_{X,Y} = 1$ y $P[aX + bY = c] = 1$ cuando $\rho_{X,Y} = -1$. De manera que en ambos casos X y Y están relacionadas linealmente. En el caso $\rho_{X,Y} = 1$, Y crece cuando X crece, mientras que en el caso $\rho_{X,Y} = -1$, Y decrece cuando X crece.

Cuando X y Y son independientes se tiene $\rho_{X,Y} = 0$. Pero el ejemplo 9.69 del primer volumen de este libro muestra que el coeficiente de correlación entre dos variables aleatorias X y Y puede ser cero sin que X y Y sean independientes.

DEFINICIÓN 2.50 (Matriz de covarianzas). Sean X_1, \dots, X_n n variables aleatorias de varianza finita. La matriz de $n \times n$ cuya componente c_{ij} (i -ésimo renglón y j -ésima columna) está dada por $c_{ij} = \text{Cov}(X_i, X_j)$ es llamada la matriz de covarianzas de X_1, \dots, X_n .

EJERCICIOS

EJERCICIO 2.1. *Se elige al azar una ficha de un juego de domino. Sean X y Y el menor y mayor, respectivamente, de los números de la ficha seleccionada. Encuentre la función de densidad de a) $X + Y$ y b) $Y - X$.*

EJERCICIO 2.2. *Se eligen, al azar y sin reemplazo, dos tarjetas de una urna que contiene 20 tarjetas numeradas del 1 al 20. Sean X y Y el menor y mayor, respectivamente, de los números de las tarjetas seleccionadas. Encuentre la función de densidad de a) $X + Y$ y b) $Y - X$.*

EJERCICIO 2.3. *Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p y sea $Z = \max(X, Y)$. Para $x \in \{0, 1, \dots\}$, encuentre $P[X = x \mid Z = x]$.*

EJERCICIO 2.4. *Sean X y Y dos variables aleatorias independientes, ambas distribuidas uniformemente en el conjunto $\{1, \dots, N\}$. Encuentre la función de densidad de:*

a) $X + Y$

b) $\min(X, Y)$

c) $\max(X, Y)$

d) $Y - X$

e) $Z = |Y - X|$

EJERCICIO 2.5. *Se tienen 2 urnas, cada una de las cuales contiene tarjetas marcadas con números entre 1 y n , de tal manera que para cada $1 \leq k \leq n$, hay k tarjetas marcadas con el número k . Se selecciona al azar una tarjeta de cada urna y se define la variable aleatoria Z como el mayor de los números seleccionados. Encuentre la distribución de Z .*

EJERCICIO 2.6. *Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo, dos bolas de una caja que contiene 12 bolas marcadas con los números $1, \dots, 12$. Sea X el mayor de los dos números de las bolas seleccionadas. Encuentre la función de densidad de X .*

EJERCICIO 2.7. *Sea X una variable aleatoria distribuida uniformemente en el conjunto $\{1, \dots, N\}$ y sea Y una variable aleatoria con distribución geométrica de parámetro p . Suponiendo que X y Y son independientes, encuentre la función de densidad de $Z = \min(X, Y)$.*

EJERCICIO 2.8. *Sean X y Y dos variables aleatorias independientes, ambas con función de densidad dada por:*

$$f(x) = \begin{cases} \frac{1}{2N} & \text{si } x \in \{1, 2, \dots, N\} \\ \frac{1}{2} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre la función de densidad de $X + Y$.

EJERCICIO 2.9. Sean X y Y dos variables aleatorias independientes, ambas con distribución binomial de parámetros (n, p) y (m, p) , respectivamente. Demuestre que $X + Y$ tiene distribución binomial con parámetros $(n + m, p)$.

EJERCICIO 2.10. Sean X y Y dos variables aleatorias independientes, ambas con distribución Poisson de parámetros λ_1 y λ_2 respectivamente. Demuestre que $X + Y$ tiene distribución Poisson de parámetro $\lambda_1 + \lambda_2$.

EJERCICIO 2.11. El número de defectos que tiene un cierto artículo tiene distribución Poisson con parámetro $\lambda = 3$. Calcule la probabilidad de encontrar más de 95 defectos en 30 artículos seleccionados al azar.

EJERCICIO 2.12. Sea X una variable aleatoria distribuida uniformemente en el conjunto $\{1, \dots, N\}$ y sea Y una variable aleatoria con función de densidad dada por:

$$f(y) = \begin{cases} \frac{1}{2} & \text{si } y = 0 \text{ ó } y = N \\ 0 & \text{en otro caso} \end{cases}$$

Asumiendo que X y Y son independientes, encuentre la función de densidad de la variable aleatoria a) $Z = X + Y$ y b) $Z = X - Y$.

EJERCICIO 2.13. Sean X y Y dos variables aleatorias independientes, ambas con función de densidad dada por:

$$f(x) = \begin{cases} \frac{2x}{N(N+1)} & \text{si } x \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre la función de densidad de la variable aleatoria $Z = X + Y$.

EJERCICIO 2.14. Sean X y Y dos variables aleatorias independientes, ambas con distribución Poisson de parámetros λ_1 y λ_2 , respectivamente. Para $z \in \{0, 1, \dots\}$ y $y \in \{0, \dots, z\}$, encuentre $P(Y = y | X + Y = z)$.

EJERCICIO 2.15. Sean X y Y dos variables aleatorias independientes, ambas con distribución binomial de parámetros n, p y m, p , respectivamente. Para $z \in \{0, \dots, n + m\}$ y $y \in \{0, \dots, \min\{z, m\}\}$, encuentre $P(Y = y | X + Y = z)$.

EJERCICIO 2.16. Sean X , Y y Z tres variables aleatorias independientes, las 3 con distribución beta de parámetros n y 1, con $n \in \mathbb{N}$. Encuentre $P[X < Y < Z]$.

EJERCICIO 2.17. Sean X y Z dos variables aleatorias independientes, ambas con distribución beta de parámetros n y 1, con $n \in \mathbb{N}$. Encuentre $P[X < Y < Z]$, en

donde Y es una variable aleatoria, independiente de X y Z , con distribución uniforme en el intervalo $(0, 1)$.

EJERCICIO 2.18. Sean X_1, X_2 y X_3 tres variables aleatorias independientes tales que X_1 y X_3 tienen distribución beta con parámetros $n, 1$ y X_2 distribución uniforme en el intervalo $(0, 1)$. ¿Cuál es la probabilidad de que X_2 quede comprendida entre X_1 y X_3 ?

EJERCICIO 2.19. Sean $\alpha, \beta \in \mathbb{N}$, Y una variable aleatoria con distribución beta de parámetros α y β y X una variable aleatoria con distribución binomial de parámetros $n = \alpha + \beta - 1$ y $p \in (0, 1)$. Demuestre que $P[Y \leq p] = P[X \geq \alpha]$.

EJERCICIO 2.20. Sean $\lambda > 0, t > 0, \alpha \in \mathbb{N}$, X_t una variable aleatoria con distribución Poisson de parámetro λt y Y una variable aleatoria con distribución gama de parámetros α y λ . Demuestre que $P[Y > t] = P[X_t \leq \alpha - 1]$. Interprete el resultado en el contexto de eventos que ocurren aleatoriamente en el tiempo, en donde λ representa el número promedio de ocurrencias por unidad de tiempo, Y el tiempo que transcurre desde el origen hasta la α -ésima ocurrencia y X_t el número de ocurrencias hasta el tiempo t .

EJERCICIO 2.21. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad de a) $Z = 2X + Y$ y b) $Z = 3X - Y$.

EJERCICIO 2.22. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre una función de densidad de $Z = 4Y - 3X$.

EJERCICIO 2.23. Un aparato utiliza un cierto componente cuyo tiempo de vida tiene distribución exponencial y dura más de 50 días con probabilidad 0.95. Cuando el componente deja de funcionar se reemplaza con otro del mismo tipo. Efectuando un sólo reemplazo, ¿cuál es la probabilidad de que el aparato se mantenga funcionando por lo menos 52 semanas?

EJERCICIO 2.24. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal, X con parámetros μ_1 y σ_1^2 , Y con parámetros μ_2 y σ_2^2 . Demuestre que la distribución de $X + Y$ es normal con parámetros $\mu_1 + \mu_2$ y $\sigma_1^2 + \sigma_2^2$.

Sugerencia: Considere primero el caso en que $\mu_1 = \mu_2 = 0$.

EJERCICIO 2.25. Sea X_1, \dots, X_n una muestra aleatoria de una distribución normal con parámetros $\mu = 4$ y $\sigma^2 = 81$ y definamos $X = \frac{1}{n} \sum_{k=1}^n X_k$. Encuentre el más pequeño número natural n tal que $P\left[\mu - \frac{1}{2} < X < \mu + \frac{1}{2}\right] \geq 0.95$.

EJERCICIO 2.26. Sea X_1, \dots, X_{25} una muestra aleatoria de una distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 9$, Y_1, Y_2, \dots, Y_{16} una muestra aleatoria de una distribución normal con parámetros $\mu = 1$ y $\sigma^2 = 16$ y definamos $X = \frac{1}{25} \sum_{k=1}^{25} X_k$, $Y = \frac{1}{16} \sum_{k=1}^{16} Y_k$. Encuentre $P[X > Y]$.

EJERCICIO 2.27. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentre la distribución de $X^2 + Y^2$.

EJERCICIO 2.28. Sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y Y una variable aleatoria con distribución exponencial de parámetro $\lambda = 1$. Asumiendo que X y Y son independientes, encuentre una función de densidad de $Z = X + Y$.

EJERCICIO 2.29. Se seleccionan, al azar y de manera independiente, dos puntos en el intervalo (a, b) . Encuentre la distribución de la distancia entre los dos puntos seleccionados.

EJERCICIO 2.30. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} c(y^2 - x^2)e^{-y} & \text{si } -y \leq x \leq y, \quad 0 < y < \infty \\ 0 & \text{en otro caso} \end{cases}$$

en donde c es una constante. Encuentre una función de densidad de a) $X + Y$ y b) $\max(X, Y)$.

EJERCICIO 2.31. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} 6(1 - x - y) & \text{si } 0 < x < 1, 0 < y < 1 - x \\ 0 & \text{en otro caso} \end{cases}$$

Demuestre que $X + Y$ tiene distribución beta.

EJERCICIO 2.32. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} c(4 - x - y) & \text{si } 0 < x < 1, 0 < y < 3, x < y < x + 2 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre el valor de c y una función de densidad de $Y - X$.

EJERCICIO 2.33. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} c(4 - x - y) & \text{si } 0 < x < 1, 0 < y < x + 2 \\ 0 & \text{en otro caso} \end{cases}$$

en donde c es una constante. Encuentre el valor de c y las funciones de densidad de X , Y y $Z = 2Y - 3X$.

EJERCICIO 2.34. Sean X y Y dos variables aleatorias continuas con función de densidad conjunta f . Encuentre una fórmula para una función de densidad de $W = aX + bY$, en donde a y b son constantes distintas de cero.

EJERCICIO 2.35. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(-1, 1)$. Encuentre las funciones de densidad de $U = 2X + Y$ y $V = 2X - Y$.

EJERCICIO 2.36. Sean X y Y dos variables aleatorias continuas con función de densidad conjunta f . Encuentre una fórmula para una función de densidad de $V = XY$.

EJERCICIO 2.37. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{6}ye^{-y} & \text{si } 0 < x < y < \infty \\ \frac{1}{3}ye^{-y} & \text{si } -\infty < -y < x < 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad de $Z = XY$.

EJERCICIO 2.38. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad de $Z = XY$.

EJERCICIO 2.39. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(-1, 1)$. Encuentre una función de densidad de $Z = XY$ y utilícela para calcular $P[-\frac{1}{4} < XY < \frac{1}{2}]$.

EJERCICIO 2.40. Sean X y Y dos variables aleatorias independientes, con funciones de densidad dadas por:

$$f_X(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & \text{si } -1 < x < 1 \\ 0 & \text{en otro caso} \end{cases} \quad y \quad f_Y(y) = \begin{cases} ye^{-y^2/2} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

respectivamente. Encuentre e identifique la distribución de $Z = XY$.

EJERCICIO 2.41. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} e^y & \text{si } 0 < x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad de $\frac{Y}{X}$.

EJERCICIO 2.42. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{6}ye^{-y} & \text{si } 0 < x < y, 0 < y < \infty \\ \frac{1}{3}ye^{-y} & \text{si } -y < x < 0, 0 < y < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad de $Z = \frac{Y}{X}$ y utilícela para calcular $P[-3 < \frac{Y}{X} < 2]$.

EJERCICIO 2.43. Sean X y Y dos variables aleatorias independientes, X con distribución exponencial de parámetro λ y Y con distribución uniforme en el intervalo $(-1, 1)$. Encuentre una función de densidad de $Z = \frac{Y}{X}$.

EJERCICIO 2.44. Sean X y Y dos variables aleatorias independientes, X con distribución gama de parámetros α y λ y Y con distribución uniforme en el intervalo $(-1, 1)$. Encuentre una función de densidad de $Z = \frac{Y}{X}$.

EJERCICIO 2.45. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α_1, λ y α_2, λ respectivamente. Encuentre una función de densidad de $Z = \frac{Y}{X}$.

EJERCICIO 2.46. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar, de parámetros $0, \sigma_X^2$ y $0, \sigma_Y^2$ respectivamente Encuentre una función de densidad de $Z = \frac{Y}{X}$.

EJERCICIO 2.47. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{2}{x^2y^2} & \text{si } 1 < x < y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre las funciones de densidad de $U = XY$ y $V = \frac{X}{Y}$.

EJERCICIO 2.48. Sea X una variable aleatoria con distribución F con n y m grados de libertad. Demuestre que $Y = \frac{1}{X}$ tiene distribución F con m y n grados de libertad.

EJERCICIO 2.49. Sea X una variable aleatoria con distribución t con n grados de libertad. Demuestre que X^2 tiene distribución F .

EJERCICIO 2.50. Sea X una variable aleatoria con distribución F con n y m grados de libertad. Demuestre que $Z = \frac{1}{1 + \frac{n}{m}X}$ tiene distribución beta.

EJERCICIO 2.51. Sea X una variable aleatoria con distribución t con n grados de libertad. Demuestre que $Y = \frac{1}{1 + \frac{1}{n}X^2}$ tiene distribución beta.

EJERCICIO 2.52. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre las distribuciones de $\min(X, Y)$ y $\max(X, Y)$.

EJERCICIO 2.53. Sean X, Y y Z tres variables aleatorias independientes, las 3 con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad de a) $X + Y + Z$ y b) $X + Y - Z$.

EJERCICIO 2.54. Sean X_1, \dots, X_n n variables aleatorias independientes, cada una con distribución exponencial de parámetro λ . Encuentre una función de densidad de a) $Y = \min(X_1, \dots, X_n)$ y b) $Z = \max(X_1, \dots, X_n)$.

EJERCICIO 2.55. Cada uno de 8 artículos tiene un tiempo de vida que se distribuye exponencialmente con parámetro λ . Los 8 artículos se ponen a funcionar simultáneamente distribuidos en dos grupos, uno formado por 5 artículos y otro formado por 3. Encuentre la probabilidad de que falle un artículo del primer grupo antes que uno del segundo.

EJERCICIO 2.56. Sean X_1, \dots, X_n n variables aleatorias independientes, cada una con distribución uniforme en el intervalo $(0, 1)$ y sea $\lambda > 0$. Demuestre que $Y = -\frac{1}{\lambda} \ln \prod_{k=1}^n X_k$ tiene distribución gama.

EJERCICIO 2.57. Sean X_1, X_2, \dots n variables aleatorias independientes, cada una con distribución exponencial de parámetro λ . Para cada $k \in \{1, 2, \dots\}$, definamos $S_k = \sum_{j=1}^k X_j$. Para $t > 0$, sea N_t el número de S_k 's que son menores o iguales que t . Encuentre la distribución de N_t .

EJERCICIO 2.58. Sea f la función de densidad beta con parámetros $\alpha_1 > 1$ y $\alpha_2 > 1$. Encuentre el punto en donde f toma su valor máximo.

EJERCICIO 2.59. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial. Encuentre una función de densidad conjunta de $X + Y$ y $Y - X$ y utilícela para encontrar $P[X + Y \leq 1, Y - X \geq 0]$.

EJERCICIO 2.60. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad conjunta de $X + Y$ y $Y - X$ y utilícela para encontrar:

$$P[X + Y \leq 1, Y - X \geq 0].$$

EJERCICIO 2.61. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} 4xy & \text{si } (0 < x < 1 \text{ y } x < y < 1) \text{ ó } (-1 < x < 0 \text{ y } x < y < 0) \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre una función de densidad conjunta de $U = X + Y$ y $V = Y - X$, grafique la región $\{(u, v) : f_{U,V}(u, v) > 0\}$ y calcule $P[U < 1, V > \frac{1}{2}]$.

EJERCICIO 2.62. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Demuestre que las variables aleatorias $U = Y - X$ y $V = \min(X, Y)$ son independientes.

EJERCICIO 2.63. Sean X y Y dos variables aleatorias independientes con distribución normal estándar. Demuestre que las variables aleatorias $U = X + Y$ y $V = \alpha X + \beta Y$ son independientes si y sólo si $\alpha + \beta = 0$.

EJERCICIO 2.64. Sean X y Y dos variables aleatorias independientes e idénticamente distribuidas tales que las variables aleatorias $U = Y - X$ y $V = \min(X, Y)$ son independientes. Asumiendo que X y Y son no negativas, absolutamente continuas y que su función de densidad común es diferenciable, demuestre que la distribución común de X y Y es exponencial.

EJERCICIO 2.65. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama, la primera con parámetros α_1 y λ , la segunda con parámetros α_2 y λ . Demuestre que las variables aleatorias Y/X y $X + Y$ son independientes.

EJERCICIO 2.66. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre una función de densidad conjunta de $U = 2X$ y $V = X + Y$.

EJERCICIO 2.67. Sean U y V dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Demuestre que las variables aleatorias $X = \sqrt{-2 \ln U} \cos 2\pi V$ y $Y = \sqrt{-2 \ln U} \sin 2\pi V$ son independientes y que ambas tienen distribución normal estándar.

EJERCICIO 2.68. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama de parámetros α y λ . a) Encuentre la función de densidad conjunta de $U = \frac{X}{X+Y}$ y $V = \frac{Y}{X+Y}$. b) ¿Existe una función de densidad conjunta de U y V ? Justifique su respuesta.

EJERCICIO 2.69. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad conjunta de $U = X + Y$ y $V = \frac{X}{X+Y}$. ¿Son U y V independientes?

EJERCICIO 2.70. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre una función de densidad conjunta de $U = Y$ y $V = \frac{X}{X+Y}$. ¿Son U y V independientes? Justifique su respuesta.

EJERCICIO 2.71. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentre una función de densidad conjunta de $U = X^2$ y $V = X^2 + Y^2$. ¿Son U y V independientes? Justifique su respuesta.

EJERCICIO 2.72. Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del círculo $x^2 + y^2 = 1$. Sean R y Θ las coordenadas polares del punto seleccionado. Encuentre una función de densidad conjunta, así como las densidades marginales de R y Θ . ¿Son R y Θ independientes? Justifique su respuesta.

EJERCICIO 2.73. Un experimento aleatorio consiste en seleccionar al azar un punto en el interior de la elipse $\frac{x^2}{9} + \frac{y^2}{4} = 1$. Sean X y Y las coordenadas cartesianas del punto seleccionado y R y Φ las coordenadas que resultan de la transformación $x = 3r \cos \theta$, $y = 2r \sin \theta$. Encuentre una función de densidad conjunta, así como las densidades marginales de R y Φ . ¿Son R y Φ independientes? Justifique su respuesta.

EJERCICIO 2.74. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre una función de densidad conjunta de $U = X$ y $V = \frac{X}{Y}$ y utilícela para calcular $P[U > \frac{1}{2}, V < 2]$.

EJERCICIO 2.75. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(-1, 1)$. Encuentre una función de densidad conjunta de $U = X$ y $V = XY$, grafique la región $\{(u, v) : f_{U,V}(u, v) > 0\}$ y calcule $P[U < \frac{1}{2}, V < \frac{1}{2}]$.

EJERCICIO 2.76. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{2}{x^2 y^2} & \text{si } x > 1, 1 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre:

a) Una función de densidad conjunta de $U = X$ y $V = XY$.

b) $P[2 < U < 4, V < 9]$

c) las funciones de densidad marginales de U y V .

EJERCICIO 2.77. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{x^2 y^2} & \text{si } x > 1, y > 1 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre:

a) Una función de densidad conjunta de $U = XY$ y $V = \frac{Y}{X}$.

b) $P[U > 2, V < 3]$

c) Las funciones de densidad marginales de U y V .

EJERCICIO 2.78. Sean X , Y y Z tres variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y,Z}(x, y, z) = \begin{cases} e^{-z} & \text{si } 0 < x < y < z \\ 0 & \text{en otro caso} \end{cases}$$

a) Demuestre que las variables aleatorias $U = X$, $V = Y - X$ y $W = Z - Y$ son independientes. b) Encuentre $P[X + Y > Z]$.

EJERCICIO 2.79. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Demuestre que $U = \min(X, Y)$ y $V = \max(X, Y) - \min(X, Y)$ son independientes.

EJERCICIO 2.80. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Demuestre que $U = \min(X, Y)$ y $V = \max(X, Y) - \min(X, Y)$ son independientes y que ambas tienen distribución exponencial.

EJERCICIO 2.81. Un sistema consiste de 2 componentes, cada uno de los cuales tiene un tiempo de vida distribuido exponencialmente con parámetro λ . Cuando un componente se acaba, inmediatamente es reemplazado por otro de las mismas características. Denotando por T_1, T_2, \dots a los tiempos entre reemplazamientos sucesivos, encuentre una función de densidad conjunta de T_1 y T_2 . ¿Son T_1 y T_2 independientes?

EJERCICIO 2.82. Un sistema consiste de 2 componentes, cada uno de los cuales tiene un tiempo de vida distribuido uniformemente en el intervalo $(0, 1)$. Cuando un componente se acaba, inmediatamente es reemplazado por otro de las mismas características. Denotando por T_1, T_2, \dots a los tiempos entre reemplazamientos sucesivos, encuentre una función de densidad conjunta de T_1 y T_2 . ¿Son T_1 y T_2 independientes?

EJERCICIO 2.83. Dado un proceso de Poisson $\{P_t : t \geq 0\}$ y números reales $0 < t_1 < t_2 < \dots < t_n$, encuentre la función de densidad conjunta de P_{t_1}, \dots, P_{t_n} .

EJERCICIO 2.84. Se le llama movimiento browniano (en honor a Robert Brown) al movimiento que presenta una pequeña partícula que se encuentra suspendida en un líquido, el cual es debido a los choques de las moléculas del líquido con la partícula. Si consideramos un movimiento browniano en una dimensión, entonces éste se puede modelar mediante una familia de variables aleatorias $\{W_t : t \geq 0\}$ de tal manera que W_t representa la posición de la partícula en el tiempo t . Norbert Wiener construyó, en el año 1922, un modelo matemático de este tipo para el movimiento browniano y, en honor a él, se define un proceso de Wiener o movimiento browniano estándar como una familia de variables aleatorias $\{W_t : t \geq 0\}$, la cual satisface las siguientes propiedades:

(i) $W_0 = 0$.

- (ii) Si $0 < t_1 < \dots < t_n$, entonces las variables aleatorias $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ son independientes.
- (iii) Las funciones $t \mapsto W_t$ son continuas.
- (iv) Si $0 \leq s < t$, entonces la variable aleatoria $W_t - W_s$ tiene distribución normal con parámetros $\mu = 0$ y $\sigma^2 = t - s$.

Dado un proceso de Wiener estándar $\{W_t : t \geq 0\}$ y n números reales $0 < t_1 < t_2 < \dots < t_n$, encuentre una función de densidad conjunta de W_{t_1}, \dots, W_{t_n} .

EJERCICIO 2.85. Se seleccionan, al azar y de manera independiente, $n + 1$ puntos en el intervalo $(0, 2n)$. Encuentre la probabilidad de que no haya dos de ellos cuya distancia sea menor que 1.

EJERCICIO 2.86. Dada una cierta producción de lámparas, se sabe que el tiempo de vida de cada una de ellas es independiente del tiempo de vida de las otras y tiene distribución exponencial con parámetro λ . Si se prenden n lámparas simultáneamente, ¿cuál es la probabilidad de que no haya alguna lámpara que deje de funcionar dentro de la hora que sigue al momento en que deja de funcionar alguna otra lámpara?

EJERCICIO 2.87. Se seleccionan, al azar y de manera independiente, n puntos en el intervalo $(0, 1)$. Si X_1, \dots, X_n son los n puntos seleccionados, encuentre la esperanza del k -ésimo estadístico de orden correspondiente a X_1, \dots, X_n .

EJERCICIO 2.88. Tres personas quedan de verse en un cierto lugar a las 10 de la mañana. Cada persona llega al lugar de la cita de manera independiente en un tiempo aleatorio distribuido uniformemente entre las 10 y las 11 de la mañana. Cada persona espera 10 minutos y, si no llega alguna otra, se va; en cambio, si se llegan a encontrar 2 de ellas, esperan a la otra hasta las 11 de la mañana. a) ¿Cuál es la probabilidad de que ningún par de personas se encuentre? b) ¿Cuál es la probabilidad de que se encuentren las 3 personas? c) ¿Cuál es el menor tiempo que debe esperar cada persona de tal manera que la probabilidad de que se encuentren las 3 sea mayor que $\frac{1}{2}$?

EJERCICIO 2.89. Tres personas quedan de verse en un cierto lugar a las 10 de la mañana. Cada persona llega al lugar de la cita de manera independiente en un tiempo aleatorio distribuido uniformemente entre las 10 y las 11 de la mañana. Cada persona espera 10 minutos y, si no llega alguna otra, se va; pero, si se llegan a encontrar 2 de ellas, esperan a la otra 10 minutos a partir del momento en que se encuentran, después de lo cual se retiran ¿Cuál es la probabilidad de que se encuentren las 3 personas?

EJERCICIO 2.90. Sean X_1, X_2 y X_3 tres variables aleatorias independientes, las 3 con distribución uniforme en el intervalo $(0, 1)$ y sean $X_{(1)}, X_{(2)}, X_{(3)}$ los estadísticos de orden correspondientes a X_1, X_2, X_3 . Encuentre $P \left[X_{(1)} < \frac{3}{4}, X_{(3)} > \frac{1}{4} \right]$.

EJERCICIO 2.91. Sean X_1 , X_2 y X_3 tres variables aleatorias independientes, las 3 con distribución exponencial de parámetro $\lambda = 1$ y sean $X_{(1)}, X_{(2)}, X_{(3)}$ los estadísticos de orden correspondientes a X_1, X_2, X_3 . Encuentre $P[X_{(1)} > \frac{1}{2}, X_{(2)} < 2]$ y $P[X_{(2)} < 1, X_{(3)} > 1]$.

EJERCICIO 2.92. Sean X_1 , X_2 y X_3 tres variables aleatorias independientes, las 3 con distribución uniforme en el intervalo $(-1, 1)$ y sean $X_{(1)}, X_{(2)}, X_{(3)}$ los estadísticos de orden correspondientes a X_1, X_2, X_3 . Encuentre $P[X_{(1)} > -\frac{1}{2}, X_{(2)} < \frac{1}{2}]$ y $P[X_{(2)} < \frac{1}{2}, X_{(3)} > \frac{1}{2}]$.

EJERCICIO 2.93. Se seleccionan, al azar y de manera independiente, tres puntos, X , Y y Z , sobre el segmento $[0, 3L]$. Encuentre la probabilidad de que la distancia entre el menor y mayor de los puntos seleccionados sea mayor que $2L$.

EJERCICIO 2.94. Encuentre la esperanza y la varianza de una variable aleatoria con distribución t con k grados de libertad.

EJERCICIO 2.95. Encuentre la esperanza y la varianza de una variable aleatoria con distribución F con n y m grados de libertad.

EJERCICIO 2.96. Encuentre la esperanza y la varianza de una variable aleatoria con distribución beta de parámetros α_1 y α_2 .

EJERCICIO 2.97. Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo, dos bolas de una caja que contiene 20 bolas marcadas con los números $1, \dots, 20$. Sea X el menor de los dos números de las bolas seleccionadas. Encuentre $E[X]$.

EJERCICIO 2.98. Un experimento aleatorio consiste en seleccionar, al azar y con reemplazo, dos bolas de una caja que contiene 20 bolas marcadas con los números $1, \dots, 20$. Sea X el mayor de los dos números de las bolas seleccionadas. Encuentre $E[X]$.

EJERCICIO 2.99. Sean X y Y dos variables aleatorias independientes, ambas distribuidas uniformemente en el conjunto $\{1, \dots, N\}$. Encuentre la esperanza de a) $U = \min(X, Y)$ y b) $V = |Y - X|$.

EJERCICIO 2.100. Sean X_1, \dots, X_n n variables aleatorias independientes, todas distribuidas uniformemente en el conjunto $\{1, \dots, N\}$. Demuestre que

$$E[\min\{X_1, \dots, X_n\}] + E[\max\{X_1, \dots, X_n\}] = N + 1$$

EJERCICIO 2.101. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre la esperanza de $Z = XY$.

EJERCICIO 2.102. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, 1)$. Encuentre el valor esperado del primer entero positivo N tal que $X_1 + \dots + X_N > 1$.

EJERCICIO 2.103. Sea X_1, X_2, \dots una sucesión de variables aleatorias absolutamente continuas, independientes e idénticamente distribuidas y:

$$A = \{\omega \in \Omega : X_1(\omega) \geq X_2(\omega) \geq \dots\}$$

Para cada $\omega \in A$, sea $N(\omega) = 0$ y, para cada $\omega \in A^c$, sea $N(\omega) \in \{2, 3, \dots\}$ tal que:

$$X_1(\omega) \geq \dots \geq X_{N(\omega)-1}(\omega) \text{ y } X_{N(\omega)-1}(\omega) < X_{N(\omega)}(\omega).$$

Encuentre $E[N]$.

EJERCICIO 2.104. Sea X_0, X_1, \dots una sucesión de variables aleatorias absolutamente continuas, independientes e idénticamente distribuidas y:

$$A = \{\omega \in \Omega : X_1(\omega) \leq X_0, X_2(\omega) \leq X_0, \dots\}$$

Para cada $\omega \in A$, sea $N(\omega) = 0$ y, para cada $\omega \in A^c$, sea $N(\omega) \in \{1, 2, \dots\}$ tal que:

$$X_1(\omega) \leq X_0(\omega), \dots, X_{N(\omega)-1}(\omega) \leq X_0(\omega) \text{ y } X_{N(\omega)}(\omega) > X_0(\omega).$$

Encuentre $E[N]$.

EJERCICIO 2.105. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $E[\max\{X_1, \dots, X_n\}]$ y b) $E[\min\{X_1, \dots, X_n\}]$.

EJERCICIO 2.106. Se eligen dos puntos, al azar y de manera independiente, en el interior de un círculo de radio 1. Encuentre el valor esperado del cuadrado de la distancia entre ellos.

EJERCICIO 2.107. En el tiempo $t = 0$, una partícula es emitida desde el origen hacia el primer cuadrante del plano de tal manera que la magnitud de su velocidad es una variable aleatoria V con distribución uniforme en el intervalo $(0, 1)$, mientras que el ángulo Θ que forma el vector velocidad con el eje x es una variable aleatoria con distribución uniforme en el intervalo $(0, \frac{\pi}{2})$. Sea X la abscisa de la posición de la partícula una unidad de tiempo después de ser lanzada. Suponiendo que V y Θ son independientes, encuentre el coeficiente de correlación entre X y V .

CAPÍTULO 3

DISTRIBUCIÓN NORMAL MULTIVARIADA

Ustedes saben que yo escribo lentamente. Esto es principalmente porque nunca estoy satisfecho hasta haber dicho tanto como sea posible en pocas palabras, y escribir brevemente toma mucho más tiempo que escribir ampliamente

Johann Carl Friedrich Gauss

3.1. Distribución normal bivariada

DEFINICIÓN 3.1 (Distribución normal bivariada). *Se dice que la pareja de variables aleatorias X y Y tiene distribución normal bivariada si existen dos variables aleatorias independientes, U y V , con distribución normal estándar, tales que $X = aU + bV + \mu$ y $Y = cU + dV + \nu$, en donde a, b, c, d, μ y ν son constantes tales que $ad - bc \neq 0$.*

Obsérvese que la condición $ad - bc \neq 0$ significa que la transformación $x = au + bv + \mu$, $y = cu + dv + \nu$ es invertible.

PROPOSICIÓN 3.2. *Supongamos que la pareja de variables aleatorias X, Y tiene distribución normal bivariada y sean μ_X, σ_X^2 y μ_Y, σ_Y^2 la esperanza y la varianza de X y Y , respectivamente, y ρ el coeficiente de correlación entre X y Y , entonces $\rho^2 \neq 1$ y una función de densidad conjunta de X, Y está dada por:*

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

Demostración

Sean U y V dos variables aleatorias independientes, con distribución normal estándar, tales que $X = aU + bV + \mu$ y $Y = cU + dV + \nu$, en donde a, b, c, d, μ y ν son constantes tales que $ad - bc \neq 0$.

Para encontrar una función de densidad conjunta de X y Y , consideremos la transformación $x = au + bv + \mu$, $y = cu + dv + \nu$, la cual tiene como inversa a $u = \frac{d}{ad-bc}(x - \mu) - \frac{b}{ad-bc}(y - \nu)$, $v = -\frac{c}{ad-bc}(x - \mu) + \frac{a}{ad-bc}(y - \nu)$, cuyo Jacobiano está dado por $\frac{1}{ad-bc}$, de manera que:

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{|ad-bc|} f_{U,V} \left(\frac{d}{ad-bc}(x - \mu) - \frac{b}{ad-bc}(y - \nu), -\frac{c}{ad-bc}(x - \mu) + \frac{a}{ad-bc}(y - \nu) \right) \\ &= \frac{1}{|ad-bc|} \frac{1}{2\pi} \exp \left\{ -\frac{c^2+d^2}{2(ad-bc)^2}(x - \mu)^2 - \frac{a^2+b^2}{2(ad-bc)^2}(y - \nu)^2 + \frac{ac+db}{(ad-bc)^2} \frac{(x-\mu)(y-\nu)}{\sigma_X \sigma_Y} \right\} \end{aligned}$$

Pero, se tiene además:

$$\mu_X = E[X] = E[aU + bV + \mu] = \mu$$

$$\mu_Y = E[Y] = E[cU + dV + \nu] = \nu$$

$$\sigma_X^2 = Var(aU + bV + \mu) = a^2 + b^2$$

$$\sigma_Y^2 = Var(cU + dV + \nu) = c^2 + d^2$$

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[(aU + bV)(cU + dV)] = ac + bd$$

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{ac+bd}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}}$$

$$1 - \rho^2 = 1 - \frac{(ac+bd)^2}{(a^2+b^2)(c^2+d^2)} = \frac{(ad-bc)^2}{(a^2+b^2)(c^2+d^2)}$$

Por lo tanto, $\rho^2 \neq 1$ y, además, la fórmula para $f_{X,Y}$ se puede escribir de la siguiente manera:

$$f_{X,Y}(x, y)$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{1}{2(1-\rho^2)} \frac{(y-\mu_Y)^2}{\sigma_Y^2} + \frac{\rho}{1-\rho^2} (x - \mu_X)(y - \mu_Y) \right\}$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

■

A continuación se presenta la gráfica de la función de densidad conjunta de un vector aleatorio (X, Y) con distribución normal bivariada.

$$\mu_X = \mu_Y = 0, \sigma_X^2 = \sigma_Y^2 = 1, \rho_{X,Y} = \frac{1}{2}$$

EJEMPLO 3.3. Sean U y V dos variables aleatorias independientes con distribución normal estándar. Definamos $X = \frac{1}{2}U - \frac{2}{3}V - \frac{5}{6}$ y $Y = -\frac{1}{2}U - \frac{1}{6}V - \frac{1}{3}$. Se tiene entonces:

$$E[X] = -\frac{5}{6}$$

$$E[Y] = -\frac{1}{3}$$

$$Var(X) = \frac{1}{4} + \frac{4}{9} = \frac{25}{36}$$

$$Var(Y) = \frac{1}{4} + \frac{1}{36} = \frac{10}{36}$$

$$Cov(X, Y) = E\left[\left(\frac{1}{2}U - \frac{2}{3}V\right)\left(-\frac{1}{2}U - \frac{1}{6}V\right)\right] = E\left[-\frac{1}{4}U^2 + \frac{1}{9}V^2\right] = -\frac{5}{36}$$

$$\rho = -\frac{1}{\sqrt{10}}$$

Así que:

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\} \\ &= \frac{6}{5\pi} \exp\left\{-\frac{1}{2}\left[\frac{8}{5}\left(x + \frac{5}{6}\right)^2 + 4\left(y + \frac{1}{3}\right)^2 + \frac{8}{5}\left(x + \frac{5}{6}\right)\left(y + \frac{1}{3}\right)\right]\right\} \end{aligned}$$

EJEMPLO 3.4. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(-1, 3)$, vector de varianzas $(4, 9)$ y coeficiente de correlación $-\frac{2}{3}$. Entonces:

$$f_{X,Y}(x, y) = \frac{1}{4\pi\sqrt{5}} \exp \left\{ -\frac{1}{2} \left[\frac{9}{20}(x+1)^2 + \frac{1}{5}(y-3)^2 + \frac{2}{5}(x+1)(y-3) \right] \right\}$$

Se tiene:

$$\frac{9}{20}x^2 + \frac{1}{5}y^2 + \frac{2}{5}xy = \frac{9}{20} \left(x + \frac{4}{9}y \right)^2 + \frac{1}{9}y^2$$

Así que:

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{1}{4\pi\sqrt{5}} \exp \left\{ -\frac{1}{2} \left[\frac{9}{20} \left((x+1) + \frac{4}{9}(y-3) \right)^2 + \frac{1}{9}(y-3)^2 \right] \right\} \\ &= \frac{1}{4\pi\sqrt{5}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{9x+4y-3}{6\sqrt{5}} \right)^2 + \left(\frac{y-3}{3} \right)^2 \right] \right\} \end{aligned}$$

Definamos:

$$U = \frac{9X+4Y-3}{6\sqrt{5}} = \frac{3\sqrt{5}}{10}X + \frac{2\sqrt{5}}{15}Y - \frac{\sqrt{5}}{10}$$

$$V = \frac{Y-3}{3} = \frac{1}{3}Y - 1$$

Entonces:

$$f_{U,V}(u, v) = \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} [u^2 + v^2] \right\}$$

Así que U y V son independientes, ambas tienen distribución normal estándar y:

$$X = \frac{2\sqrt{5}}{3}U - \frac{4}{3}V - 1$$

$$Y = 3V + 3$$

DEFINICIÓN 3.5 (**Formas cuadráticas bidimensionales**). Se dice que una función $F : \mathbb{R}^2 \mapsto \mathbb{R}$ es una forma cuadrática si tiene la forma $F(x, y) = ax^2 + bxy + cy^2$, en donde a , b y c son constantes.

DEFINICIÓN 3.6 (**Formas cuadráticas bidimensionales definidas positivas**). Se dice que una forma cuadrática $F : \mathbb{R}^2 \mapsto \mathbb{R}$ es definida positiva si $F(x, y) > 0$ para cualquier vector $(x, y) \neq (0, 0)$.

PROPOSICIÓN 3.7. Una forma cuadrática $F(x, y) = ax^2 + bxy + cy^2$ es definida positiva si y sólo si $a > 0$, $c > 0$ y $4ac - b^2 > 0$.

Demostración

Si $a \neq 0$, se tiene:

$$F(x, y) = a \left[\left(x + \frac{b}{2a}y \right)^2 + \frac{4ac-b^2}{4a^2}y^2 \right]$$

así que, cuando $a > 0$ y $4ac - b^2 > 0$, la forma cuadrática F es definida positiva.

Por otra parte, si F es definida positiva, obsérvese primero que no es posible tener $a = 0$ y $c = 0$ pues en ese caso se tendría $F(x, y) = bxy$, lo cual no define una forma cuadrática definida positiva. Además, $a \neq 0$ pues de otra forma se tendría:

$$F(x, y) = bxy + cy^2 = c \left[\left(y + \frac{b}{2c}x \right)^2 - \frac{b^2}{4c^2}x^2 \right]$$

lo cual define una forma cuadrática que no es definida positiva.

Así que se tiene:

$$F(x, y) = a \left[\left(x + \frac{b}{2a}y \right)^2 + \frac{4ac-b^2}{4a^2}y^2 \right]$$

y, como F es definida positiva, necesariamente se tiene $a > 0$ y $4ac - b^2 > 0$. Finalmente, $c > \frac{b^2}{4a} \geq 0$. ■

Si el vector aleatorio (X, Y) tiene distribución normal bivariada, entonces su función de densidad conjunta $f_{X,Y}$ tiene la forma $f_{X,Y}(x, y) = K \exp \left\{ -\frac{1}{2} [F(x - \mu, y - \nu)] \right\}$, en donde F es la forma cuadrática definida por:

$$F(x, y) = \frac{1}{(1-\rho^2)} \left[\frac{1}{\sigma_X^2}x^2 - \frac{2\rho}{\sigma_X\sigma_Y}xy + \frac{1}{\sigma_Y^2}y^2 \right] = \frac{1}{(1-\rho^2)\sigma_X^2}x^2 + \frac{1}{(1-\rho^2)\sigma_Y^2}y^2 - \frac{2\rho}{(1-\rho^2)\sigma_X\sigma_Y}xy$$

Esta forma cuadrática es definida positiva ya que:

$$\frac{1}{(1-\rho^2)\sigma_X^2} > 0$$

$$\frac{1}{(1-\rho^2)^2} \left[\frac{4\rho^2}{\sigma_X^2\sigma_Y^2} - \frac{4}{\sigma_X^2\sigma_Y^2} \right] = -\frac{4}{(1-\rho^2)\sigma_X^2\sigma_Y^2} < 0$$

PROPOSICIÓN 3.8. *Sea X, Y una pareja de variables aleatorias con función de densidad conjunta $f_{X,Y}$ dada por $f_{X,Y}(x, y) = K \exp \left\{ -\frac{1}{2} [F(x - \mu, y - \nu)] \right\}$, en donde F es una forma cuadrática definida positiva y K, μ y ν son constantes, entonces el vector (X, Y) tiene distribución normal bivariada.*

Demostración

Sea $F(x, y) = ax^2 + bxy + cy^2$, entonces $F(x, y) = a \left(x + \frac{b}{2a}y \right)^2 + \frac{4ac-b^2}{4a}y^2$, así que:

$$f_{X,Y}(x, y) = K \exp \left\{ -\frac{1}{2} \left[a \left((x - \mu) + \frac{b}{2a}(y - \nu) \right)^2 + \frac{4ac-b^2}{4a}(y - \nu)^2 \right] \right\}$$

Consideremos entonces las variables aleatorias:

$$U = \sqrt{a} \left((X - \mu) + \frac{b}{2a}(Y - \nu) \right)$$

$$V = \sqrt{\frac{4ac - b^2}{4a}}(Y - \nu)$$

La transformación que define a la pareja U, V en términos de la pareja X, Y es invertible, así que existen constantes A, B, C y D tales que $AD - BC \neq 0$ y $X = AU + BV + \mu$, $Y = CU + DV + \nu$. Además, se tiene:

$$f_{U,V}(u, v) = C \exp \left\{ -\frac{1}{2} [u^2 + v^2] \right\}$$

en donde C es una constante.

Por lo tanto, U y V son independientes y ambas tienen distribución normal estándar. Así que la pareja X, Y tiene distribución normal bivariada. ■

Sea $f : \mathbb{R}^2 \mapsto \mathbb{R}$ una función de densidad de la forma:

$$f(x, y) = C \exp \left\{ -\frac{1}{2} (ax^2 + bxy + cy^2 + dx + ey) \right\}$$

en donde d, e y C son constantes y la forma cuadrática $F(x, y) = ax^2 + bxy + cy^2$ es definida positiva, es decir, por la proposición 3.7, $a > 0$, $c > 0$ y $4ac - b^2 > 0$.

Para μ y ν números reales cualesquiera, se tiene:

$$\begin{aligned} F(x - \mu, y - \nu) &= a(x - \mu)^2 + b(x - \mu)(y - \nu) + c(y - \nu)^2 \\ &= ax^2 + bxy + cy^2 - (b\nu + 2a\mu)x - (2c\nu + b\mu)y + a\mu^2 + b\mu\nu + c\nu^2 \end{aligned}$$

Como $b^2 - 4ac \neq 0$, el sistema de ecuaciones:

$$b\nu + 2a\mu = -d$$

$$2c\nu + b\mu = -e$$

tiene una única solución para μ y ν . Así que f puede escribirse en la forma siguiente:

$$\begin{aligned} f(x, y) &= K \exp \left\{ -\frac{1}{2} [F(x - \mu, y - \nu)] \right\} \\ &= K \exp \left\{ -\frac{1}{2} [a(x - \mu)^2 + b(x - \mu)(y - \nu) + c(y - \nu)^2] \right\} \end{aligned}$$

en donde K es una constante.

Por lo tanto, f es función de densidad de una distribución normal bivariada.

Sea (X, Y) un vector aleatorio con esa función de densidad. Entonces f puede escribirse también en la forma siguiente:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu)^2}{\sigma_X^2} + \frac{(y-\nu)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu)(y-\nu)}{\sigma_X\sigma_Y} \right] \right\}$$

en donde σ_X^2 y σ_Y^2 son las varianzas de X y Y , respectivamente, y ρ es el coeficiente de correlación entre X y Y .

Se tiene entonces:

$$(1 - \rho^2) \sigma_X^2 = \frac{1}{a}$$

$$(1 - \rho^2) \sigma_Y^2 = \frac{1}{c}$$

$$\frac{2\rho}{\sigma_X\sigma_Y(1-\rho^2)} = -b$$

Así que:

$$\rho = -\frac{b}{\sqrt{4ac}}$$

$$\sigma_X^2 = \frac{1}{a} \frac{1}{1-\rho^2}$$

$$\sigma_Y^2 = \frac{1}{c} \frac{1}{1-\rho^2}$$

$$Cov(X, Y) = \rho\sigma_X\sigma_Y = -\frac{b}{4ac-b^2}$$

En particular, se tiene el siguiente resultado:

PROPOSICIÓN 3.9. *Sea $f : \mathbb{R}^2 \mapsto \mathbb{R}$ una función de densidad de la forma:*

$$f(x, y) = C \exp \left\{ -\frac{1}{2} (ax^2 + bxy + cy^2 + dx + ey) \right\}$$

en donde a, b, c, d, e y C son constantes y la forma cuadrática F definida por $F(x, y) = ax^2 + bxy + cy^2$ es definida positiva. Entonces f es función de densidad de una distribución normal bivariada.

EJEMPLO 3.10. *Sea X, Y una pareja de variables aleatorias con función de densidad conjunta $f_{X,Y}$ dada por:*

$$f_{X,Y}(x, y) = C \exp \left\{ -\frac{1}{2} (2x^2 - 2xy + 2y^2 - 4x) \right\}$$

en donde C es una constante.

Como la forma cuadrática $F(x, y) = 2x^2 - 2xy + 2y^2 = 2 \left(x - \frac{1}{2}y \right)^2 + \frac{3}{2}y^2$ es definida positiva, la distribución conjunta del vector aleatorio (X, Y) es normal bivariada.

Sean μ y ν las esperanzas de X y Y , respectivamente, entonces:

$$\begin{aligned} F(x - \mu, y - \nu) &= 2(x - \mu)^2 - 2(x - \mu)(y - \nu) + 2(y - \nu)^2 \\ &= 2x^2 - 2xy + 2y^2 - 4\left(\mu - \frac{1}{2}\nu\right)x + (2\mu - 4\nu)y + 2\mu^2 - 2\mu\nu + 2\nu^2 \end{aligned}$$

Así que:

$$\mu - \frac{1}{2}\nu = 1$$

$$2\mu - 4\nu = 0$$

Por lo tanto, $\mu = \frac{4}{3}$ y $\nu = \frac{2}{3}$.

Así que:

$$f_{X,Y}(x, y) = K \exp \left\{ -\frac{1}{2} \left[2 \left(x - \frac{4}{3} \right)^2 - 2 \left(x - \frac{4}{3} \right) \left(y - \frac{2}{3} \right) + 2 \left(y - \frac{2}{3} \right)^2 \right] \right\}$$

en donde K es una constante.

Sean σ_X^2 y σ_Y^2 las varianzas de X y Y , respectivamente, y ρ el coeficiente de correlación entre X y Y . Se tiene entonces:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu)^2}{\sigma_X^2} + \frac{(y-\nu)^2}{\sigma_Y^2} - 2\rho\frac{(x-\mu)(y-\nu)}{\sigma_X\sigma_Y} \right] \right\}$$

Así que:

$$(1 - \rho^2) \sigma_X^2 = \frac{1}{2}$$

$$(1 - \rho^2) \sigma_Y^2 = \frac{1}{2}$$

$$\frac{2\rho}{\sigma_X\sigma_Y(1-\rho^2)} = 2$$

Por lo tanto:

$$\rho = \frac{1}{2}$$

$$\sigma_x^2 = \frac{2}{3}$$

$$\sigma_Y^2 = \frac{2}{3}$$

$$\text{Cov}(X, Y) = \frac{1}{3}$$

$$K = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} = \frac{\sqrt{3}}{2\pi}$$

Es decir:

$$f_{X,Y}(x,y) = \frac{\sqrt{3}}{2\pi} \exp \left\{ -\frac{1}{2} \left[2 \left(x - \frac{4}{3} \right)^2 - 2 \left(x - \frac{4}{3} \right) \left(y - \frac{2}{3} \right) + 2 \left(y - \frac{2}{3} \right)^2 \right] \right\}$$

▲

Evidentemente, si la pareja de variables aleatorias X, Y tiene distribución normal bivariada y su coeficiente de correlación es ρ , entonces el vector aleatorio (X, Y) es absolutamente continuo, X y Y tienen distribución normal y $\rho^2 \neq 1$. Sin embargo, debe de observarse que la aseveración inversa no es válida, es decir, para que X, Y tenga distribución normal bivariada no basta con que el vector aleatorio (X, Y) sea absolutamente continuo, que X y Y tengan distribución normal y que su coeficiente de correlación ρ satisfaga $\rho^2 \neq 1$. En efecto, considérese el ejemplo siguiente:

EJEMPLO 3.11. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} e^{-\frac{1}{2}(x^2+y^2)} & \text{si } x \geq 0, y < 0 \text{ ó } x < 0, y \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

La pareja X, Y no tiene distribución normal bivariada pues, si la tuviera, su función de densidad conjunta no se anularía en ningún punto. Sin embargo, se tiene:

$$f_X(x) = \begin{cases} \frac{1}{\pi} \int_0^\infty e^{-\frac{1}{2}(x^2+y^2)} dy & \text{si } x < 0 \\ \frac{1}{\pi} \int_{-\infty}^0 e^{-\frac{1}{2}(x^2+y^2)} dy & \text{si } x \geq 0 \end{cases} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$f_Y(y) = \begin{cases} \frac{1}{\pi} \int_0^\infty e^{-\frac{1}{2}(x^2+y^2)} dx & \text{si } y < 0 \\ \frac{1}{\pi} \int_{-\infty}^0 e^{-\frac{1}{2}(x^2+y^2)} dx & \text{si } y \geq 0 \end{cases} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

Así que tanto X como Y tienen distribución normal estándar. Además:

$$\begin{aligned} Cov(X, Y) &= E[XY] = \frac{1}{\pi} \int_{-\infty}^0 \int_0^\infty xye^{-\frac{1}{2}(x^2+y^2)} dy dx + \frac{1}{\pi} \int_0^\infty \int_{-\infty}^0 xye^{-\frac{1}{2}(x^2+y^2)} dy dx \\ &= \frac{1}{\pi} \int_{-\infty}^0 xe^{-\frac{1}{2}x^2} dx - \frac{1}{\pi} \int_0^\infty xe^{-\frac{1}{2}x^2} dx = -\frac{2}{\pi} \end{aligned}$$

Así que, $\rho_{X,Y}^2 = \frac{4}{\pi^2} \neq 1$.

▲

Una propiedad importante de una pareja de variables aleatorias con distribución normal bivariada consiste en que basta con que su coeficiente de correlación sea cero para poder asegurar que tales variables aleatorias son independientes. Este resultado se obtiene inmediatamente de la fórmula:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \exp \left\{ -\frac{1}{2(1-\rho_{X,Y}^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho_{X,Y} \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

3.2. Un poco de Cálculo Matricial

Asumimos que el lector está familiarizado con la definición y las operaciones básicas de suma y producto entre matrices de números real o por un número real α :

$$\alpha \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1m} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \cdots & \alpha a_{nm} \end{pmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nm} + b_{nm} \end{pmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nr} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rm} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^r a_{1j}b_{j1} & \sum_{j=1}^r a_{1j}b_{j2} & \cdots & \sum_{j=1}^r a_{1j}b_{jm} \\ \sum_{j=1}^r a_{2j}b_{j1} & \sum_{j=1}^r a_{2j}b_{j2} & \cdots & \sum_{j=1}^r a_{2j}b_{jm} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^r a_{nj}b_{j1} & \sum_{j=1}^r a_{nj}b_{j2} & \cdots & \sum_{j=1}^r a_{nj}b_{jm} \end{pmatrix}$$

También asumimos que el lector está familiarizado con las propiedades del determinante de una matriz A de $n \times n$, el cual será denotado por $|A|$.

Si para cada $i \in \{1, \dots, n\}$ y $j \in \{1, \dots, m\}$, A_{ij} es una matriz, la notación:

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{pmatrix}$$

representa a la matriz que se obtiene al escribir los componentes de cada una de las matrices A_{ij} . Esta notación se utilizará únicamente cuando las matrices que pertenecen al mismo renglón tengan el mismo número de renglones y las matrices que pertenecen a la misma columna tengan el mismo número de columnas. Por ejemplo, si:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix}$$

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{pmatrix}$$

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \\ d_{41} & d_{42} & d_{43} \end{pmatrix}$$

entonces, la notación $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ representa a la matriz:

$$\begin{pmatrix} a_{11} & a_{12} & b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & b_{21} & b_{22} & b_{23} \\ c_{11} & c_{12} & d_{11} & d_{12} & d_{13} \\ c_{21} & c_{22} & d_{21} & d_{22} & d_{23} \\ c_{31} & c_{32} & d_{31} & d_{32} & d_{33} \\ c_{41} & c_{42} & d_{41} & d_{42} & d_{43} \end{pmatrix}$$

Esta notación es cómoda sobre todo porque, para fines de las operaciones entre matrices, se puede operar con las matrices que forman un arreglo matricial como si se tratara de las entradas de una matriz usual; obviamente, esto únicamente cuando las operaciones que se realicen estén bien definidas. Por ejemplo, si:

$$A' = \begin{pmatrix} a'_{11} & a'_{12} \\ a'_{21} & a'_{22} \end{pmatrix}$$

$$B' = \begin{pmatrix} b'_{11} & b'_{12} & b'_{13} \\ b'_{21} & b'_{22} & b'_{23} \end{pmatrix}$$

$$C' = \begin{pmatrix} e'_{11} & e'_{12} & e'_{13} & e'_{14} \\ e'_{21} & e'_{22} & e'_{23} & e'_{24} \end{pmatrix}$$

$$D' = \begin{pmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \\ c'_{31} & c'_{32} \end{pmatrix}$$

$$E' = \begin{pmatrix} d'_{11} & d'_{12} & d'_{13} \\ d'_{21} & d'_{22} & d'_{23} \\ d'_{31} & d'_{32} & d'_{33} \end{pmatrix}$$

$$F' = \begin{pmatrix} f'_{11} & f'_{12} & f'_{13} & f'_{14} \\ f'_{21} & f'_{22} & f'_{23} & f'_{24} \\ f'_{31} & f'_{32} & f'_{33} & f'_{34} \end{pmatrix}$$

entonces:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} A' & B' & C' \\ D' & E' & F' \end{pmatrix} = \begin{pmatrix} AA' + BD' & AB' + BE' & AC' + BF' \\ CA' + DD' & CB' + DE' & CC' + DF' \end{pmatrix}$$

Una matriz A con n renglones y m columnas, es decir, de $n \times m$, representa una transformación lineal de \mathbb{R}^m en \mathbb{R}^n . En efecto, representando a los vectores mediante matrices de una columna, la transformación que asocia a cada vector $x \in \mathbb{R}^m$ el vector $y = Ax \in \mathbb{R}^n$, es lineal. En forma desarrollada, si a_{ij} son los elementos de la matriz A , x_1, \dots, x_m las coordenadas de x y y_1, \dots, y_n las coordenadas de y , entonces, para $i \in \{1, \dots, n\}$, se tiene $y_i = \sum_{j=1}^m a_{ij}x_j$.

DEFINICIÓN 3.12 (Matrices diagonales). Diremos que una matriz (a_{ij}) de $n \times n$ es diagonal si $a_{ij} = 0$ para cualquier pareja $i, j \in \{1, \dots, n\}$ tal que $i \neq j$.

Dados m números reales d_1, \dots, d_m , denotaremos por $\mathbf{D}_{d_1, \dots, d_m}$ a la matriz diagonal, de $m \times m$, con entradas d_1, \dots, d_m , es decir:

$$D_{d_1, d_2, \dots, d_m} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_m \end{pmatrix}$$

DEFINICIÓN 3.13 (Matriz identidad). *La matriz diagonal de $n \times n$ para la cual $a_{ii} = 1$, para cualquier $i \in \{1, \dots, n\}$, será llamada la matriz identidad de $n \times n$ y será denotada por I_n .*

DEFINICIÓN 3.14 (Matrices triangulares superiores). *Se dice que una matriz de $n \times n$ es triangular superior (resp. inferior) si todos sus elementos que están debajo (resp. arriba) de la diagonal son 0.*

La matriz de $n \times m$ formada exclusivamente por ceros será denotada por $\mathbf{0}_{nm}$.

DEFINICIÓN 3.15 (Matrices invertibles). *Diremos que una matriz A de $n \times n$ es invertible, si existe una matriz, denotada por A^{-1} , tal que $AA^{-1} = A^{-1}A = I_n$.*

Obviamente, si A es invertible, entonces A^{-1} también lo es y $(A^{-1})^{-1} = A$.

El siguiente es uno de los resultados básicos del Álgebra Lineal, su demostración puede consultarse, por ejemplo, en Grossman, S. I., Álgebra Lineal con aplicaciones, McGraw-Hill.

PROPOSICIÓN 3.16. *Si A una matriz de $n \times n$, las siguientes condiciones son equivalentes:*

- a) A es invertible.
- b) Para cada $b \in \mathbb{R}^n$ existe un único vector $x \in \mathbb{R}^n$ tal que $Ax = b$.
- c) No existe ningún vector distinto de cero $x \in \mathbb{R}^n$ tal que $Ax = 0$.
- d) El determinante de A es distinto de cero.

COROLARIO 3.17. *Sea A una matriz de $n \times n$ y supongamos que existe una matriz B tal que $BA = I_n$. Entonces A es invertible y $A^{-1} = B$.*

Demostración

Sea $x \in \mathbb{R}^n$ tal que $Ax = 0$, entonces $x = I_n x = BAx = 0$; así que, por la proposición 3.16, A es invertible. Además:

$$B = I_n B = B(AA^{-1}) = (BA)A^{-1} = I_n A^{-1} = A^{-1}$$

■

COROLARIO 3.18. *Sea A una matriz de $n \times n$ y supongamos que existe una matriz B tal que $AB = I_n$. Entonces A es invertible y $A^{-1} = B$.*

Demostración

Por el corolario 3.17, B es invertible y $B^{-1} = A$, así que A también es invertible y $A^{-1} = (B^{-1})^{-1} = B$.

■

Un método para determinar si una matriz es invertible y, en su caso, encontrar su inversa, consiste en transformar la matriz en una cuyos elementos sean exclusivamente 0's y 1's mediante la aplicación repetida de las siguientes operaciones: a) multiplicación de los elementos de un renglón de la matriz por un número real distinto de cero, b) adición de los elementos de un renglón de la matriz a los de otro renglón de la misma y c) intercambio de los elementos de dos renglones de la matriz. Este proceso siempre permite transformar la matriz original en una con las siguientes propiedades: a) los renglones formados exclusivamente por 0's son los últimos, b) el primer elemento distinto de cero de cada renglón es un 1, c) el primer elemento distinto de 0 en cualquier renglón se encuentra a la derecha del primer elemento distinto de 0 del renglón anterior y d) la columna en donde se encuentra el primer elemento distinto de 0 de un renglón tiene exclusivamente 0's en sus otras entradas. La forma de la matriz que se obtiene es conocida como **escalonada reducida** y el método para obtenerla es llamado de Gauss-Jordan.

Por ejemplo, las siguientes matrices están en su forma escalonada reducida:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 6 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 4 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 3 & 0 & 8 & 0 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Se puede demostrar que una matriz A de $n \times n$ es invertible si y sólo si su forma escalonada reducida es la identidad y la inversa de A es la matriz que se obtiene al aplicar a la matriz identidad exactamente las mismas operaciones, y en el mismo orden que se efectuaron sobre A , para obtener su forma escalonada reducida.

Por ejemplo, consideremos la siguiente matriz:

$$A = \begin{pmatrix} 0 & -1 & \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & 1 & 1 & 0 & 0 \\ -\frac{1}{2} & -1 & 0 & 1 & 0 \\ 1 & 2 & 1 & -2 & \frac{1}{2} \\ 0 & -\frac{1}{2} & 1 & 0 & -1 \end{pmatrix}$$

Vamos a aplicar el método de Gauss-Jordan para llevar A a su forma escalonada reducida, aplicando, simultáneamente, las mismas operaciones sobre A y la matriz identidad:

$$\begin{aligned} & \begin{pmatrix} 0 & -1 & \frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{2} & 1 & 1 & 0 & 0 \\ -\frac{1}{2} & -1 & 0 & 1 & 0 \\ 1 & 2 & 1 & -2 & \frac{1}{2} \\ 0 & -\frac{1}{2} & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\ \rightarrow & \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & -2 & 1 & 2 & -1 \\ -1 & -2 & 0 & 2 & 0 \\ 2 & 4 & 2 & -4 & 1 \\ 0 & -1 & 2 & 0 & -2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} \\ \rightarrow & \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & -2 & -4 & 1 \\ 0 & -1 & 2 & 0 & -2 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 \\ 0 & -4 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} \\ \rightarrow & \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & -2 & -4 & 1 \\ 0 & 0 & \frac{3}{2} & -1 & -\frac{3}{2} \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & -4 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 2 \end{pmatrix} \\ \rightarrow & \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & -\frac{5}{2} & -\frac{3}{2} \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ -1 & -\frac{3}{2} & -\frac{3}{2} & 0 & 2 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&\rightarrow \begin{pmatrix} 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ \frac{4}{11} & -\frac{4}{11} & \frac{16}{11} & \frac{10}{11} & -\frac{8}{11} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{28}{11} & -\frac{6}{11} & -\frac{20}{11} & \frac{4}{11} & -\frac{12}{11} \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ \frac{4}{11} & -\frac{4}{11} & \frac{16}{11} & \frac{10}{11} & -\frac{8}{11} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{28}{11} & -\frac{6}{11} & -\frac{20}{11} & \frac{4}{11} & -\frac{12}{11} \\ -\frac{12}{11} & \frac{12}{11} & -\frac{4}{11} & -\frac{8}{11} & \frac{2}{11} \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 \\ \frac{4}{11} & -\frac{4}{11} & \frac{16}{11} & \frac{10}{11} & -\frac{8}{11} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{28}{11} & -\frac{6}{11} & -\frac{20}{11} & \frac{4}{11} & -\frac{12}{11} \\ -\frac{12}{11} & \frac{12}{11} & -\frac{4}{11} & -\frac{8}{11} & \frac{2}{11} \\ -\frac{2}{11} & \frac{14}{11} & \frac{6}{11} & \frac{4}{11} & \\ \frac{2}{11} & \frac{11}{9} & \frac{11}{3} & \frac{11}{6} & -\frac{4}{11} \\ \frac{4}{11} & -\frac{4}{11} & \frac{16}{11} & \frac{10}{11} & -\frac{8}{11} \end{pmatrix}
\end{aligned}$$

Así que A es invertible y:

$$A^{-1} = \begin{pmatrix} \frac{28}{11} & -\frac{6}{11} & -\frac{20}{11} & \frac{4}{11} & -\frac{12}{11} \\ -\frac{12}{11} & \frac{12}{11} & -\frac{4}{11} & -\frac{8}{11} & \frac{2}{11} \\ -\frac{2}{11} & \frac{14}{11} & \frac{6}{11} & \frac{4}{11} & \\ \frac{2}{11} & \frac{11}{9} & \frac{11}{3} & \frac{11}{6} & -\frac{4}{11} \\ \frac{4}{11} & -\frac{4}{11} & \frac{16}{11} & \frac{10}{11} & -\frac{8}{11} \end{pmatrix} = \frac{1}{11} \begin{pmatrix} 28 & -6 & -20 & 4 & -12 \\ -12 & 12 & -4 & -8 & 2 \\ -2 & 2 & 14 & 6 & 4 \\ 2 & 9 & -3 & -6 & -4 \\ 4 & -4 & 16 & 10 & -8 \end{pmatrix}$$

DEFINICIÓN 3.19 (Matrices transpuestas). *La transpuesta de una matriz A de $n \times m$ es una matriz de $m \times n$ la cual se obtiene colocando los renglones de A como columnas. Esta matriz será denotada por A^t .*

Se puede demostrar que el determinante de una matriz de $n \times n$ es igual al determinante de su transpuesta, así que la matriz es invertible si y sólo si lo es su transpuesta.

Recuérdese que un vector $x \in \mathbb{R}^n$ se representa mediante una matriz de una columna. Por lo tanto x^t es una matriz de un renglón. Además, $\|x\|^2 = x^t x$.

DEFINICIÓN 3.20 (Matrices simétricas). Diremos que una matriz A es simétrica si $A^t = A$.

DEFINICIÓN 3.21 (Matrices ortogonales). Diremos que una matriz A , de $n \times n$, es ortogonal si $A^t A = I_n$.

Por el Corolario 3.17, si A es ortogonal, entonces A es invertible y $A^{-1} = A^t$. De aquí se sigue a su vez que A es ortogonal si y sólo si su transpuesta es ortogonal.

DEFINICIÓN 3.22 (Formas cuadráticas n-dimensionales). Se dice que una función $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una forma cuadrática si tiene la forma:

$$F(x_1, \dots, x_n) = \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j$$

en donde los coeficientes a_{ij} son constantes.

DEFINICIÓN 3.23 (Formas cuadráticas n-dimensionales definidas positivas). Se dice que una forma cuadrática $F : \mathbb{R}^n \mapsto \mathbb{R}$ es definida positiva si $F(x) > 0$ para cualquier vector $x \neq 0$.

Una manera, conocida como el **método de Lagrange**, para investigar si una forma cuadrática es definida positiva consiste en ir completando cuadrados para expresarla como suma de cuadrados.

EJEMPLO 3.24. Consideremos la forma cuadrática $F : \mathbb{R}^5 \mapsto \mathbb{R}$ definida por:

$$F(x_1, x_2, x_3, x_4, x_5) = x_1^2 + 2x_2^2 + 3x_4^2 + x_5^2 + 2x_1x_2 - x_1x_3 - 2x_1x_4 + x_2x_4 - 2x_3x_5$$

Lo primero que podemos observar es que los coeficientes de los términos al cuadrado son no negativos; si alguno de ellos fuera negativo, la forma cuadrática no sería definida positiva pues, haciendo cero las otras coordenadas, podríamos encontrar un vector distinto de cero para el cual la forma cuadrática sería negativa.

En seguida podemos analizar cómo es la forma cuadrática en cada pareja de variables cuyo producto aparezca en la forma cuadrática; en este caso tenemos las siguientes:

$$F_1(x_1, x_2) = x_1^2 + 2x_2^2 + 2x_1x_2 = (x_1 + x_2)^2 + x_2^2$$

$$F_2(x_1, x_3) = x_1^2 - x_1x_3 = \left(x_1 - \frac{1}{2}x_3\right)^2 - \frac{1}{4}x_3^2$$

$$F_3(x_1, x_4) = x_1^2 + 3x_4^2 - 2x_1x_4 = (x_1 - x_4)^2 + 2x_4^2$$

$$F_4(x_2, x_4) = 2x_2^2 + 3x_4^2 + x_2x_4 = 2\left(x_2 + \frac{1}{4}x_4\right)^2 + \frac{23}{8}x_4^2$$

$$F_5(x_3, x_5) = x_5^2 - 2x_3x_5 = (x_5 - x_3)^2 - x_3^2$$

De estas cinco formas cuadráticas, F_1 , F_3 y F_4 son definidas positivas ya que son no negativas y se hacen cero únicamente cuando las dos variables son nulas. En cambio, F_2 y F_5 no son definidas positivas pues, por ejemplo, para cualquier valor distinto de cero de x_3 , tomando $x_1 = \frac{1}{2}x_3$ y $x_5 = x_3$, se tiene $F_2(x_1, x_3) = -\frac{1}{4}x_3^2 < 0$ y $F_5(x_3, x_5) = -x_3^2 < 0$.

Con esto es suficiente para concluir que la forma cuadrática F no es definida positiva pues si lo fuera, sería definida positiva en cualquier subconjunto de variables, haciendo nulas las otras.

EJEMPLO 3.25. Consideremos la forma cuadrática $F : \mathbb{R}^5 \mapsto \mathbb{R}$ definida por:

$$F(x_1, x_2, x_3, x_4, x_5) = x_1^2 + 2x_2^2 + x_3^2 + 10x_4^2 + 5x_5^2 \\ + 2x_1x_2 - x_1x_3 - 2x_1x_4 + x_2x_4 + 5x_3x_4 + 2x_3x_5 + 10x_4x_5$$

En este caso se tiene:

$$F_1(x_1, x_2) = x_1^2 + 2x_2^2 + 2x_1x_2 = (x_1 + x_2)^2 + x_2^2$$

$$F_2(x_1, x_3) = x_1^2 + x_3^2 - x_1x_3 = \left(x_1 - \frac{1}{2}x_3\right)^2 + \frac{3}{4}x_3^2$$

$$F_3(x_1, x_4) = x_1^2 + 10x_4^2 - 2x_1x_4 = (x_1 - x_4)^2 + 9x_4^2$$

$$F_4(x_2, x_4) = 2x_2^2 + 10x_4^2 + x_2x_4 = 2\left(x_2 + \frac{1}{4}x_4\right)^2 + \frac{79}{8}x_4^2$$

$$F_5(x_3, x_4) = x_3^2 + 10x_4^2 + 5x_3x_4 = \left(x_3 + \frac{5}{2}x_4\right)^2 + \frac{15}{4}x_4^2$$

$$F_6(x_3, x_5) = x_3^2 + 5x_5^2 + 2x_3x_5 = (x_3 + x_5)^2 + 4x_5^2$$

$$F_7(x_4, x_5) = 10x_4^2 + 5x_5^2 + 10x_4x_5 = 10\left(x_4 + \frac{1}{2}x_5\right)^2 + \frac{5}{2}x_5^2$$

Las siete formas cuadráticas que se obtienen con cada pareja de variables cuyo producto aparece en la forma cuadrática, son definidas positivas. Esto no contradice el que F sea definida positiva pero no es suficiente para mostrar que lo es.

Expresemos entonces F , completa, como suma de cuadrados, lo cual se hace completando cuadrados en cada una de las variables:

$$F(x_1, x_2, x_3, x_4, x_5) = (x_1^2 + 2x_1x_2 - x_1x_3 - 2x_1x_4) \\ + 2x_2^2 + x_3^2 + 10x_4^2 + 5x_5^2 + x_2x_4 + 5x_3x_4 + 2x_3x_5 + 10x_4x_5 \\ = \left(x_1 + x_2 - \frac{1}{2}x_3 - x_4\right)^2 - \left(x_2 - \frac{1}{2}x_3 - x_4\right)$$

$$\begin{aligned}
& +2x_2^2 + x_3^2 + 10x_4^2 + 5x_5^2 + x_2x_4 + 5x_3x_4 + 2x_3x_5 + 10x_4x_5 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2^2 + x_2x_3 + 3x_2x_4) \\
& + \frac{3}{4}x_3^2 + 4x_3x_4 + 9x_4^2 + 5x_5^2 + 2x_3x_5 + 10x_4x_5 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 - (\frac{1}{2}x_3 + \frac{3}{2}x_4)^2 \\
& + \frac{3}{4}x_3^2 + 9x_4^2 + 5x_5^2 + 4x_3x_4 + 2x_3x_5 + 10x_4x_5 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 + \frac{1}{2}(x_3^2 + 5x_3x_4 + 4x_3x_5) \\
& + \frac{27}{4}x_4^2 + 5x_5^2 + 10x_4x_5 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 \\
& + \frac{1}{2}(x_3 + \frac{5}{2}x_4 + 2x_5)^2 - \frac{1}{2}(\frac{5}{2}x_4 + 2x_5)^2 \\
& + \frac{27}{4}x_4^2 + 5x_5^2 + 10x_4x_5 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 + \frac{1}{2}(x_3 + \frac{5}{2}x_4 + 2x_5)^2 \\
& + \frac{29}{8}(x_4^2 + \frac{40}{29}x_4x_5) + 3x_5^2 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 + \frac{1}{2}(x_3 + \frac{5}{2}x_4 + 2x_5)^2 \\
& + \frac{29}{8}(x_4 + \frac{20}{29}x_5)^2 - \frac{29}{8}(\frac{20}{29}x_5)^2 + 3x_5^2 \\
& = (x_1 + x_2 - \frac{1}{2}x_3 - x_4)^2 + (x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_4)^2 + \frac{1}{2}(x_3 + \frac{5}{2}x_4 + 2x_5)^2 \\
& + \frac{29}{8}(x_4 + \frac{20}{29}x_5)^2 + \frac{37}{29}x_5^2
\end{aligned}$$

Así que F es definida positiva pues es no negativa y se hace cero únicamente cuando las 5 variables son nulas.

▲

Si $F(x_1, x_2, \dots, x_n) = \sum_{\{i,j \in \{1,2,\dots,n\}: i \leq j\}} c_{ij}x_i x_j$ es una forma cuadrática definida positiva entonces $c_{ii} > 0$ y $4c_{ii}c_{jj} - c_{ij}^2 > 0$ para cualquier pareja i, j . En efecto, consideremos la forma cuadrática $F_{ij} : \mathbb{R}^2 \mapsto \mathbb{R}$ definida por:

$$F_{ij}(x_i, x_j) = c_{ii}x_i^2 + c_{ij}x_i x_j + c_{jj}x_j^2$$

la cual es definida positiva, así que, por la proposición 3.7, se tiene:

$c_{ii} > 0$, $c_{jj} > 0$ y $4c_{ii}c_{jj} - c_{ij}^2 > 0$

Iniciando el método de Lagrange, de completación de cuadrados, se obtiene:

$$\begin{aligned}
F(x_1, \dots, x_n) &= \left(\sqrt{c_{11}}x_1 + \frac{c_{12}}{2\sqrt{c_{11}}}x_2 + \dots + \frac{c_{1n}}{2\sqrt{c_{11}}}x_n \right)^2 \\
&- \left(\frac{c_{12}}{2\sqrt{c_{11}}}x_2 + \dots + \frac{c_{1n}}{2\sqrt{c_{11}}}x_n \right)^2 + \sum_{i=2}^n c_{ii}x_i^2 + \sum_{\{i,j \in \{2, \dots, n\}: i < j\}} c_{ij}x_i x_j \\
&= \left(\sqrt{c_{11}}x_1 + \frac{c_{12}}{2\sqrt{c_{11}}}x_2 + \dots + \frac{c_{1n}}{2\sqrt{c_{11}}}x_n \right)^2 - \sum_{i=2}^n \frac{c_{1i}^2}{4c_{11}}x_i^2 - 2 \sum_{\{i,j \in \{2, \dots, n\}: i < j\}} \frac{c_{1i}c_{1j}}{4c_{11}}x_i x_j \\
&+ \sum_{i=2}^n c_{ii}x_i^2 + \sum_{\{i,j \in \{2, \dots, n\}: i < j\}} c_{ij}x_i x_j \\
&= \left(\sqrt{c_{11}}x_1 + \frac{c_{12}}{2\sqrt{c_{11}}}x_2 + \dots + \frac{c_{1n}}{2\sqrt{c_{11}}}x_n \right)^2 \\
&+ \left(\frac{4c_{11}c_{22} - c_{12}^2}{4c_{11}} \right) x_2^2 + \sum_{\{j \in \{3, \dots, n\}\}} \left(\frac{2c_{11}c_{2j} - c_{12}c_{1j}}{2c_{11}} \right) x_2 x_j \\
&+ \sum_{i=3}^n c_{ii}^2 x_i^2 + \sum_{\{i,j \in \{3, \dots, n\}: i < j\}} c_{ij}x_i x_j - 2 \sum_{\{i,j \in \{3, \dots, n\}: i < j\}} \frac{c_{1i}c_{1j}}{4c_{11}}x_i x_j \\
&= \left(\sqrt{c_{11}}x_1 + \frac{c_{12}}{2\sqrt{c_{11}}}x_2 + \dots + \frac{c_{1n}}{2\sqrt{c_{11}}}x_n \right)^2 \\
&+ \left(\frac{\sqrt{4c_{11}c_{22} - c_{12}^2}}{2\sqrt{c_{11}}}x_2 + \frac{1}{2} \frac{\sqrt{c_{11}}}{\sqrt{4c_{11}c_{22} - c_{12}^2}} \frac{2c_{11}c_{23} - c_{12}c_{13}}{c_{11}}x_3 + \dots + \frac{1}{2} \frac{\sqrt{c_{11}}}{\sqrt{4c_{11}c_{22} - c_{12}^2}} \frac{2c_{11}c_{2n} - c_{12}c_{1n}}{c_{11}}x_n \right)^2 \\
&- \left(\frac{1}{2} \frac{\sqrt{c_{11}}}{\sqrt{4c_{11}c_{22} - c_{12}^2}} \frac{2c_{11}c_{23} - c_{12}c_{13}}{c_{11}}x_3 + \dots + \frac{1}{2} \frac{\sqrt{c_{11}}}{\sqrt{4c_{11}c_{22} - c_{12}^2}} \frac{2c_{11}c_{2n} - c_{12}c_{1n}}{c_{11}}x_n \right)^2 \\
&+ \sum_{i=3}^n c_{ii}^2 x_i^2 + \sum_{\{i,j \in \{3, \dots, n\}: i < j\}} c_{ij}x_i x_j - 2 \sum_{\{i,j \in \{3, \dots, n\}: i < j\}} \frac{c_{1i}c_{1j}}{4c_{11}}x_i x_j \\
&= (a_{11}x_1 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 + \dots
\end{aligned}$$

en donde a_{11} y a_{22} son positivos.

PROPOSICIÓN 3.26. *Una forma cuadrática $F(x_1, \dots, x_n) = \sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} c_{ij}x_i x_j$ es definida positiva si y sólo si el método de Lagrange, de completación de cuadrados, puede continuarse hasta obtener una expresión de la forma siguiente:*

$$\begin{aligned}
F(x_1, \dots, x_n) &= (a_{11}x_1 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\
&+ \dots + (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 + (a_{nn}x_n)^2
\end{aligned}$$

en donde $a_{jj} > 0$ para cualquier $j \in \{1, \dots, n\}$.

Demstración

Supongamos primero que el método de Lagrange, de completación de cuadrados, puede continuarse hasta obtener una expresión de la forma:

$$F(x_1, \dots, x_n) = (a_{11}x_1 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\ + \dots + (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 + (a_{nn}x_n)^2$$

en donde $a_{jj} > 0$ para cualquier $j \in \{1, \dots, n\}$.

Se tiene $F(x_1, \dots, x_n) \geq 0$ para cualquier vector $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Además, si $F(x_1, \dots, x_n) = 0$ entonces todos los términos de la sumatoria son 0, es decir:

$$a_{nn}x_n = 0$$

$$a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n = 0$$

⋮

$$a_{22}x_2 + \dots + a_{2n}x_n = 0$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0$$

Así que entonces, $x_n = x_{n-1} = \dots = x_2 = x_1 = 0$.

Por lo tanto, F es definida positiva.

Supongamos ahora que F es definida positiva.

Primero demostraremos que, completando cuadrados, F puede llevarse siempre a la forma siguiente:

$$F(x_1, \dots, x_n) = (a_{11}x_1 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\ \pm (a_{33}x_3 + \dots + a_{3n}x_n)^2 + \dots \pm (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 \pm (a_{nn}x_n)^2$$

en donde $a_{jj} > 0$ para cualquier $j \in \{1, \dots, n\}$.

En efecto, supongamos que después de $m - 1$ cuadrados ya no hay términos conteniendo x_m^2 , entonces consideremos la forma cuadrática que se obtiene de F aplicándola a un vector cuyas coordenadas, después de la m -ésima, son cero. Esta nueva forma cuadrática, $F_m(x_1, \dots, x_m)$, sigue siendo definida positiva y se tiene:

$$F_m(x_1, \dots, x_m) = (a_{11}x_1 + \dots + a_{1m}x_m)^2 + (a_{22}x_2 + \dots + a_{2m}x_m)^2 \\ \pm \dots \pm (a_{(m-)(m-1)}x_{m-1} + a_{(m-1)m}x_m)^2$$

Consideremos un vector con coordenada $x_m = 1$. Las primeras $m - 1$ coordenadas de ese vector pueden tomarse de tal forma que los m cuadrados de la sumatoria sean 0, así que se tendría $F(x_1, \dots, x_{m-1}, 1) = 0$, lo cual es una contradicción. Por lo tanto, el proceso de completación de cuadrados continúa hasta obtenerse una expresión de la forma:

$$F(x_1, \dots, x_n) = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\ \pm \dots \pm (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 \pm (a_{nn}x_n)^2$$

en donde $a_{jj} \neq 0$ para cualquier $j \in \{1, \dots, n\}$.

a_{jj} puede hacerse positivo ya que se encuentra dentro de una expresión al cuadrado.

Ahora demostraremos que todos los signos de la suma de cuadrados son positivos. En efecto, supongamos que el m -ésimo término de esta suma de cuadrados es el primero con signo negativo, entonces consideremos la forma cuadrática que se obtiene de F aplicándola a un vector cuyas coordenadas, después de la m -ésima, son cero. Esta nueva forma cuadrática, $F_m(x_1, \dots, x_m)$, sigue siendo definida positiva y se tiene:

$$F_m(x_1, \dots, x_m) = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m)^2 + (a_{22}x_2 + \dots + a_{2m}x_m)^2 \\ + \dots + (a_{(m-)(m-1)}x_{m-1} + a_{(m-1)m}x_m)^2 - (a_{mm}x_m)^2$$

Consideremos un vector con coordenada $x_m = 1$. Las primeras $m - 1$ coordenadas de ese vector pueden tomarse de tal forma que los primeros $m - 1$ cuadrados de la sumatoria sean 0, así que se tendría $F(x_1, \dots, x_{m-1}, 1) \leq 0$, lo cual es una contradicción. Por lo tanto, no existe ningún término con signo negativo en la sumatoria.

Se tiene entonces:

$$F(x_1, \dots, x_n) = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\ + \dots + (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 + (a_{nn}x_n)^2$$

en donde $a_{jj} > 0$ para cualquier $j \in \{1, \dots, n\}$.

■

DEFINICIÓN 3.27 (Matriz asociada a una forma cuadrática). Si $F : \mathbb{R}^n \mapsto \mathbb{R}$ es una forma cuadrática dada por:

$$F(x_1, \dots, x_n) = \sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} x_i x_j,$$

$$\text{entonces la matriz } Q \text{ definida por } Q = \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

es llamada la matriz asociada a la forma cuadrática F .

Obsérvese que la matriz Q asociada a una forma cuadrática $F : \mathbb{R}^n \mapsto \mathbb{R}$ es simétrica y satisface la relación $F(x) = (Qx) \cdot x = x^t Qx$

Obsérvese también que si Q es una matriz simétrica arbitraria, entonces la función $F : \mathbb{R}^n \mapsto \mathbb{R}$ definida por $F(x) = x^t Qx$ es una forma cuadrática.

DEFINICIÓN 3.28 (Matrices definidas positivas). *Se dice que una matriz simétrica de $n \times n$ es definida positiva si su forma cuadrática asociada es definida positiva.*

PROPOSICIÓN 3.29. *Sea A una matriz de $n \times n$ invertible, entonces la matriz $Q = AA^t$ es simétrica y definida positiva*

Demostración

$Q^t = (AA^t)^t = AA^t = Q$, así que Q es simétrica.

$$x^t Qx = x^t AA^t x = (A^t x)^t (A^t x) = (A^t x) \cdot (A^t x) \geq 0$$

Además, como A es invertible, $A^t x = 0$ si y sólo si $x = 0$. ■

COROLARIO 3.30. *Sea A una matriz de $n \times n$ invertible, entonces la matriz $Q = A^t A$ es simétrica y definida positiva.*

PROPOSICIÓN 3.31. *Sea A una matriz de $n \times n$ tal que la matriz $Q = A^t A$ es simétrica y definida positiva, entonces A es invertible.*

Demostración

Sea F la forma cuadrática definida por Q y $x \in \mathbb{R}^n$ distinto de 0, entonces:

$$\|Ax\|^2 = (Ax) \cdot (Ax) = (Ax)^t (Ax) = x^t A^t Ax = x^t Qx = F(x) > 0$$

Por lo tanto, $Ax \neq 0$.

Es decir, no existe ningún vector $x \in \mathbb{R}^n$, distinto de 0, tal que $Ax = 0$. Así que, por la proposición 3.16, A es invertible. ■

COROLARIO 3.32. *Sea A una matriz de $n \times n$ tal que la matriz $Q = AA^t$ es simétrica y definida positiva, entonces A es invertible.*

Combinando las proposiciones 3.29 y 3.31, así como sus corolarios, se tienen los siguientes resultados:

PROPOSICIÓN 3.33. *Sea A una matriz de $n \times n$, entonces la matriz $Q = AA^t$ es simétrica y definida positiva si y sólo si A es invertible.*

COROLARIO 3.34. *Sea A una matriz de $n \times n$, entonces la matriz $Q = A^tA$ es simétrica y definida positiva si y sólo si A es invertible.*

Por otra parte, la proposición 3.26 nos lleva al siguiente resultado:

PROPOSICIÓN 3.35. *Sea Q una matriz simétrica de $n \times n$, entonces Q es definida positiva si y sólo si existe una matriz invertible B triangular superior tal que $B^tB = Q$.*

Demostración

Si existe una matriz invertible B triangular superior tal que $B^tB = Q$, entonces de la proposición 3.29 se sigue que Q es definida positiva.

Supongamos ahora que Q es definida positiva y consideremos la forma cuadrática que define:

$$F(x_1, \dots, x_n) = x^t Q x = \sum_{i=1}^n c_{ii}^2 x_i^2 + \sum_{\{i,j \in \{1, \dots, n\} : i < j\}} 2c_{ij} x_i x_j$$

en donde $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

De acuerdo con la proposición 3.26, F puede expresarse de la siguiente manera:

$$F(x_1, \dots, x_n) = (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)^2 + (a_{22}x_2 + \dots + a_{2n}x_n)^2 \\ + \dots + (a_{(n-1)(n-1)}x_{n-1} + a_{(n-1)n}x_n)^2 + (a_{nn}x_n)^2$$

en donde $a_{jj} > 0$ para cualquier $j \in \{1, \dots, n\}$.

$$\text{Sea } B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & \dots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix}$$

entonces:

$$F(x_1, \dots, x_n) = (Bx) \cdot (Bx) = x^t (B^t B) x$$

Por lo tanto $B^t B = Q$, la cual es una matriz simétrica y definida positiva, así que B es invertible. ■

Combinando este último resultado con los anteriores, se tiene la siguiente proposición:

PROPOSICIÓN 3.36. *Una matriz simétrica Q es definida positiva si y sólo si existe una matriz A invertible tal que $Q = A^t A$.*

COROLARIO 3.37. *Una matriz simétrica Q es definida positiva si y sólo si existe una matriz A invertible tal que $Q = AA^t$.*

COROLARIO 3.38. *Sea Q una matriz simétrica y definida positiva. Entonces Q es invertible y su inversa es simétrica y definida positiva*

Demostración

Por la proposición 3.36, existe una matriz invertible A tal que $Q = A^t A$, así que Q es invertible. Además, $Q^{-1} = A^{-1} (A^t)^{-1} = A^{-1} (A^{-1})^t$, así que, por la proposición 3.29, Q^{-1} es simétrica y definida positiva. ■

Recordemos que la forma cuadrática F que define una matriz simétrica Q está dada por $F(x) = x^t Q x$. Así que si $Q = A^t A$, entonces:

$$F(x) = x^t A^t A x = (Ax) \cdot (Ax) = \|Ax\|^2$$

Es decir, se tienen los siguientes resultados:

PROPOSICIÓN 3.39. *Una forma cuadrática $F : \mathbb{R}^n \mapsto \mathbb{R}$ es definida positiva si y sólo si existe una matriz invertible A de $n \times n$ tal que $F(x) = \|Ax\|^2$ para todo $x \in \mathbb{R}^n$.*

PROPOSICIÓN 3.40. *Sea A una matriz de $n \times n$, entonces la forma cuadrática $F : \mathbb{R}^n \mapsto \mathbb{R}$ definida por $F(x) = \|Ax\|^2$ es definida positiva si y sólo si A es invertible.*

3.2.1. Valores y vectores propios de matrices simétricas.

PROPOSICIÓN 3.41. *Sea A una matriz simétrica de $n \times n$, entonces existe un número real α y un vector $x \in \mathbb{R}^n$, de norma 1, tal que $Ax = \alpha x$.*

Demostración

Por el teorema fundamental del álgebra, el polinomio en la variable compleja z , $P(z) = |A - zI_n|$, tiene por lo menos una raíz, es decir, existe un número complejo $z = \alpha + \beta i$

tal que $|A - zI| = 0$. El conjugado de z , $\bar{z} = \alpha - \beta i$ es entonces también raíz del mismo polinomio. Sea $B = (A - zI)(A - \bar{z}I) = A^2 - 2\alpha A + (\alpha^2 + \beta^2)I$, entonces, como el determinante de B es nulo, existe un vector distinto de cero $y \in \mathbb{R}^n$ tal que $By = 0$. Entonces, definiendo $x = \frac{y}{\|y\|}$, se tiene $Bx = 0$, $\|x\| = 1$ y:

$$\begin{aligned} 0 &= x^t Bx = x^t A^2 x - 2\alpha x^t A x + (\alpha^2 + \beta^2) x^t x \\ &= x^t A^t A x - \alpha x^t A^t x - \alpha x^t A x + \alpha^2 + \beta^2 \\ &= (x^t A^t - \alpha x^t)(Ax - \alpha x) + \beta^2 \\ &= (Ax - \alpha x)^t (Ax - \alpha x) + \beta^2 \\ &= \|Ax - \alpha x\|^2 + \beta^2 \end{aligned}$$

Así que, $\beta = 0$ y $Ax - \alpha x = 0$.

■

PROPOSICIÓN 3.42. *Sea Q una matriz simétrica de $n \times n$, entonces existe un número real α y una matriz ortogonal P tal que $P^t Q P$ tiene la forma:*

$$P^t Q P = \begin{pmatrix} \alpha & 0_{1(n-1)} \\ 0_{(n-1)1} & Q' \end{pmatrix}$$

en donde Q' es una matriz de $(n-1) \times (n-1)$ simétrica.

Demostración

Sea $\alpha \in \mathbb{R}$ y $x \in \mathbb{R}^n$, de norma 1, tal que $Qx = \alpha x$.

Sean x_1, \dots, x_n las coordenadas de x . Como $x \neq 0$, por lo menos una de sus coordenadas es distinta de cero; supongamos $x_j \neq 0$. Entonces, para $k \in \{1, \dots, j-1\}$, sea w_k el vector de \mathbb{R}^n cuya k -ésima coordenada es 1 y todas las demás son cero y, para $k \in \{j, \dots, n-1\}$, sea w_k el vector de \mathbb{R}^n cuya $(k+1)$ -ésima coordenada es 1 y todas las demás son cero.

Obviamente, los vectores w_1, \dots, w_{n-1} son linealmente independientes. Además, si $\lambda_0, \dots, \lambda_{n-1}$ son tales que $\lambda_0 x + \dots + \lambda_{n-1} w_{n-1} = 0$ entonces, $\lambda_0 x_j = 0$, así que $\lambda_0 = 0$. Por lo tanto, como w_1, \dots, w_{n-1} son linealmente independientes, también se tiene $\lambda_1 = \dots = \lambda_{n-1} = 0$. Es decir, los vectores x, w_1, \dots, w_{n-1} son linealmente independientes.

Siguiendo el procedimiento de ortogonalización de Gram-Schmidt, definamos $v_1 = x$ y, para $k \in \{2, \dots, n\}$:

$$v_k = w_{k-1} - \sum_{i=1}^{k-1} \frac{1}{\|v_i\|^2} (v_i \cdot w_{k-1}) v_i$$

Entonces $v_1 \neq 0$ y si $v_2 = w_1 - (v_1 \cdot w_1)v_1 = 0$, x y w_1 serían linealmente dependientes, lo cual es una contradicción. Por lo tanto, $v_2 \neq 0$.

Además:

$$v_1 \cdot v_2 = v_1 \cdot (w_1 - (v_1 \cdot w_1)v_1) = (v_1 \cdot w_1) - (v_1 \cdot w_1) = 0$$

Así que v_1 y v_2 son ortogonales.

Supongamos ahora que v_1, \dots, v_m son todos distintos de cero y ortogonales por parejas. Entonces $v_{m+1} = w_m - \sum_{i=1}^m \frac{1}{\|v_i\|^2} (v_i \cdot w_m)v_i$ está bien definido y si $v_{m+1} = 0$, los vectores x, w_1, \dots, w_m serían linealmente dependientes, lo cual es una contradicción. Por lo tanto, $v_{m+1} \neq 0$.

Además, si $k \in \{1, \dots, m\}$, se tiene:

$$\begin{aligned} v_k \cdot v_{m+1} &= v_k \cdot \left(w_m - \sum_{i=1}^m \frac{1}{\|v_i\|^2} (v_i \cdot w_m)v_i \right) \\ &= (v_k \cdot w_m) - v_k \cdot \left(\sum_{i=1}^m \frac{1}{\|v_i\|^2} (v_i \cdot w_m)v_i \right) \\ &= (v_k \cdot w_m) - \sum_{i=1}^m \frac{1}{\|v_i\|^2} (v_i \cdot w_m)(v_i \cdot v_k) \\ &= (v_k \cdot w_m) - \frac{1}{\|v_k\|^2} (v_k \cdot w_m)(v_k \cdot v_k) = 0 \end{aligned}$$

Así que v_k y v_{m+1} son ortogonales.

Por lo tanto, v_1, \dots, v_n están bien definidos, todos son distintos de cero y son ortogonales.

Para $k \in \{1, \dots, n\}$, sea $u_k = \frac{v_k}{\|v_k\|}$, entonces los vectores u_1, \dots, u_n son ortogonales y de norma 1.

Sea P la matriz cuyas columnas son los vectores u_1, \dots, u_n . Obviamente, P es una matriz ortogonal y su primera columna está formada por las coordenadas de x , así que tiene la forma siguiente:

$$P = \begin{pmatrix} x_1 & p_{12} & \cdots & p_{1n} \\ x_2 & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_n & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

Además, como $Qx = \alpha x$, P^tQP tiene la forma siguiente:

$$P^tQP = \begin{pmatrix} \alpha & s_{12} & \cdots & s_{1n} \\ 0 & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & s_{n2} & \cdots & s_n \end{pmatrix}$$

Pero $(P^tQP)^t = P^tQ^tP = P^tQP$, así que P^tQP es simétrica. Por lo tanto, tiene la forma:

$$P^tQP = \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & s_{n2} & \cdots & s_n \end{pmatrix}$$

y la matriz:

$$Q' = \begin{pmatrix} s_{22} & \cdots & s_{2n} \\ \vdots & \ddots & \vdots \\ s_{n2} & \cdots & s_n \end{pmatrix}$$

es simétrica. ■

PROPOSICIÓN 3.43. *Sea Q una matriz simétrica de $n \times n$, entonces existen n números reales $\alpha_1, \dots, \alpha_n$ y una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \dots, \alpha_n}$.*

Demostración

Por la proposición 3.42, existe $\alpha_1 \in \mathbb{R}$ y una matriz ortogonal P_1 tal que $P_1^tQP_1$ tiene la forma:

$$P_1^tQP_1 = \begin{pmatrix} \alpha_1 & 0_{1(n-1)} \\ 0_{(n-1)1} & Q_1 \end{pmatrix}$$

en donde Q_1 es una matriz de $(n-1) \times (n-1)$ simétrica.

Consideremos ahora $k \in \{1, \dots, n-2\}$ y supongamos que existen k números reales $\alpha_1, \dots, \alpha_k$ y una matriz ortogonal P_k tal que $P_k^tQP_k$ tiene la forma:

$$P_k^tQP_k = \begin{pmatrix} D_{\alpha_1, \dots, \alpha_k} & 0_{k(n-k)} \\ 0_{(n-k)k} & Q_k \end{pmatrix}$$

en donde Q_k es una matriz de $(n-k) \times (n-k)$ simétrica.

Nuevamente por la proposición 3.42, existe $\alpha_{k+1} \in \mathbb{R}$ y una matriz ortogonal R , de $(n-k) \times (n-k)$, tal que $R^t Q_k R$ tiene la forma:

$$R^t Q_k R = \begin{pmatrix} \alpha_{k+1} & 0_{1(n-k-1)} \\ 0_{(n-k-1)1} & Q_{k+1} \end{pmatrix}$$

en donde Q_{k+1} es una matriz de $(n-k-1) \times (n-k-1)$ simétrica.

Definamos:

$$S = \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R \end{pmatrix}$$

Entonces:

$$\begin{aligned} S^t S &= \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R^t \end{pmatrix} \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R \end{pmatrix} \\ &= \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R^t R \end{pmatrix} = \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & I_{n-k} \end{pmatrix} = I_n \end{aligned}$$

Así que S es ortogonal y, por lo tanto, $P_{k+1} = P_k S$ también es ortogonal. Además:

$$\begin{aligned} P_{k+1}^t Q P_{k+1} &= S^t P_k^t Q P_k S \\ &= \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R^t \end{pmatrix} \begin{pmatrix} D_{\alpha_1, \dots, \alpha_k} & 0_{k(n-k)} \\ 0_{(n-k)k} & Q_k \end{pmatrix} \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R \end{pmatrix} \\ &= \begin{pmatrix} I_k & 0_{k(n-k)} \\ 0_{(n-k)k} & R^t \end{pmatrix} \begin{pmatrix} D_{\alpha_1, \dots, \alpha_k} & 0_{k(n-k)} \\ 0_{(n-k)k} & Q_k R \end{pmatrix} \\ &= \begin{pmatrix} D_{\alpha_1, \dots, \alpha_k} & 0_{k(n-k)} \\ 0_{(n-k)k} & R^t Q_k R \end{pmatrix} = \begin{pmatrix} D_{\alpha_1, \dots, \alpha_{k+1}} & 0_{(k+1)(n-k-1)} \\ 0_{(n-k-1)(k+1)} & Q_{k+1} \end{pmatrix} \end{aligned}$$

Así que, para cualquier $m \in \{1, \dots, n-1\}$, existen m números reales $\alpha_1, \dots, \alpha_m$ y una matriz ortogonal P_m tal que $P_m^t Q P_m$ tiene la forma:

$$P_m^t Q P_m = \begin{pmatrix} D_{\alpha_1, \dots, \alpha_m} & 0_{m(n-m)} \\ 0_{(n-m)m} & Q_m \end{pmatrix}$$

en donde Q_m es una matriz de $(n-m) \times (n-m)$ simétrica.

En particular, para $m = n-1$, existen $n-1$ números reales $\alpha_1, \dots, \alpha_{n-1}$ y una matriz ortogonal P tal que $P^t Q P$ tiene la forma:

$$P^tQP = \begin{pmatrix} D_{\alpha_1, \alpha_2, \dots, \alpha_{n-1}} & 0_{(n-1)1} \\ 0_{1(n-1)} & Q_{n-1} \end{pmatrix}$$

en donde Q_{n-1} es una matriz de 1×1 simétrica, es decir, Q_{n-1} tiene un único elemento α_n . Así que, $P^tQP = D_{\alpha_1, \dots, \alpha_n}$. ■

Sea Q una matriz simétrica de $n \times n$, y P una matriz ortogonal tal que $P^tQP = D_{\alpha_1, \dots, \alpha_n}$, en donde $\alpha_1, \dots, \alpha_n$ son números reales. Se tiene entonces $QP = PD_{\alpha_1, \dots, \alpha_n}$. Sea x_k el vector que forma la k -ésima columna de P . Entonces Qx_k forma la k -ésima columna de $PD_{\alpha_1, \dots, \alpha_n}$; pero, la k -ésima columna de $PD_{\alpha_1, \dots, \alpha_n}$ está dada por el vector $\alpha_k x_k$. Por lo tanto, se tiene, $Qx_k = \alpha_k x_k$ y, como $x_k \neq 0$, el determinante $|Q - \alpha_k I_n|$ es nulo. Es decir, $\alpha_1, \dots, \alpha_n$ son raíces del polinomio en z , $|Q - zI_n|$ y, para cada $k \in \{1, \dots, n\}$, la k -ésima columna de P está formada por un vector x_k de norma 1 tal que $Qx_k = \alpha_k x_k$. En otras palabras, se tiene el siguiente resultado:

COROLARIO 3.44. *Sea Q una matriz simétrica de $n \times n$, entonces existen n números reales $\alpha_1, \dots, \alpha_n$ y n vectores de norma 1, x_1, \dots, x_n , ortogonales entre sí, tales que $\alpha_1, \dots, \alpha_n$ son raíces del polinomio $p(z) = |Q - zI|$ y, para $i \in \{1, \dots, n\}$, $Q(x_i) = \alpha_i x_i$. Además, si P es la matriz cuyas columnas están formadas por los vectores x_1, \dots, x_n , entonces P es ortogonal y satisface la relación $P^tQP = D_{\alpha_1, \dots, \alpha_n}$.*

En la terminología del Álgebra Lineal, si Q es una matriz de $n \times n$, el polinomio $p(z) = |Q - zI_n|$ es llamado el **polinomio característico** de Q . También, si α es un número real o complejo para el cual existe un vector $x \in \mathbb{R}^n$ tal que $Q(x) = \alpha x$, entonces α es llamado un **valor propio** de Q y se dice que x es un **vector propio** correspondiente a α .

También se tiene el siguiente resultado:

PROPOSICIÓN 3.45. *Sean α y β dos valores propios distintos de la matriz simétrica Q y sean x, y vectores propios correspondientes a α y β , respectivamente, entonces x y y son ortogonales.*

Demostración

Como $\alpha \neq \beta$, por lo menos uno de ellos es distinto de 0. Supongamos $\alpha \neq 0$. Entonces:

$$\begin{aligned} \alpha\beta x \cdot y &= (\alpha x) \cdot (\beta y) = (Qx) \cdot (Qy) = y^t Q^t Qx = y^t Q^2 x \\ &= y^t Q(Qx) = y^t Q(\alpha x) = \alpha y^t Q(x) = \alpha^2 y^t x = \alpha^2 x \cdot y \end{aligned}$$

Así que:

$$\alpha(\beta - \alpha)x \cdot y = 0$$

Por lo tanto, $x \cdot y = 0$, así que x y y son ortogonales.

■

EJEMPLO 3.46. Sea $Q = \begin{pmatrix} 9 & 5 & 3 & 3 \\ 5 & -7 & -5 & -5 \\ 3 & -5 & 6 & 0 \\ 3 & -5 & 0 & 6 \end{pmatrix}$

El determinante $|Q - zI_4|$ está dado por:

$$p(z) = z^4 - 14z^3 - 96z^2 + 2016z - 6912$$

Las raíces de este polinomio son $\rho_1 = -12$, $\rho_2 = 6$, $\rho_3 = 8$ y $\rho_4 = 12$.

Sea ρ una de las raíces de p . Para obtener un vector propio asociado a ρ , se tiene que resolver el sistema de ecuaciones $Qx = \rho x$ cuya matriz está dada por:

$$\begin{pmatrix} 9 - \rho & 5 & 3 & 3 \\ 5 & -7 - \rho & -5 & -5 \\ 3 & -5 & 6 - \rho & 0 \\ 3 & -5 & 0 & 6 - \rho \end{pmatrix}$$

la cual se puede llevar a su forma escalonada reducida de la siguiente manera:

$$\begin{aligned} &\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 9 - \rho & 5 & 3 & 3 \\ 3 & -5 & 6 - \rho & 0 \\ 3 & -5 & 0 & 6 - \rho \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & -\rho^2 + 2\rho + 88 & -5\rho + 60 & -5\rho + 60 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 3\rho - 4 & 15 & -5\rho + 45 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & -\rho^2 + 2\rho + 88 & -5\rho + 60 & -5\rho + 60 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \rho - 6 & -\rho + 6 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \rho - 6 & -\rho + 6 \\ 0 & -\rho^2 + 2\rho + 88 & -5\rho + 60 & -5\rho + 60 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \rho - 6 & -\rho + 6 \\ 0 & 0 & \frac{-\rho^3 + 8\rho^2 + 110\rho - 840}{3\rho - 4} & 2\frac{17\rho - 156}{3\rho - 4} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \frac{-\rho^3 + 8\rho^2 + 110\rho - 840}{3\rho - 4} & \frac{34\rho - 312}{3\rho - 4} \\ 0 & 0 & \rho - 6 & -(\rho - 6) \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 5 & -\rho - 7 & -5 & -5 \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \frac{-\rho^3 + 8\rho^2 + 110\rho - 840}{3\rho - 4} & \frac{34\rho - 312}{3\rho - 4} \\ 0 & 0 & 0 & \frac{(\rho + 12)(\rho - 6)(\rho - 8)(\rho - 12)}{-\rho^3 + 8\rho^2 + 110\rho - 840} \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & \frac{-\rho^2 - \rho + 67}{3\rho - 4} & \frac{25}{3\rho - 4} \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & \frac{-\rho^3 + 8\rho^2 + 110\rho - 840}{3\rho - 4} & \frac{34\rho - 312}{3\rho - 4} \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & \frac{-\rho^2 - \rho + 67}{3\rho - 4} & \frac{25}{3\rho - 4} \\ 0 & 3\rho - 4 & -5\rho + 45 & 15 \\ 0 & 0 & 1 & \frac{34\rho - 312}{-\rho^3 + 8\rho^2 + 110\rho - 840} \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&\rightarrow \begin{pmatrix} 1 & 0 & 0 & \frac{(\rho - 6)(3\rho - 4)}{-\rho^3 + 8\rho^2 + 110\rho - 840} \\ 0 & 1 & 0 & -\frac{5(\rho - 6)(\rho - 12)}{-\rho^3 + 8\rho^2 + 110\rho - 840} \\ 0 & 0 & 1 & \frac{34\rho - 312}{-\rho^3 + 8\rho^2 + 110\rho - 840} \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

Por lo tanto, una solución está dada por el vector:

$$w = \begin{pmatrix} -(\rho - 6)(3\rho - 4) \\ 5(\rho - 6)(\rho - 12) \\ -34\rho + 312 \\ -\rho^3 + 8\rho^2 + 110\rho - 840 \end{pmatrix}$$

Así que 4 vectores propios, correspondientes a los valores propios -12 , 6 , 8 y 12 , están dados, respectivamente, por:

$$\begin{pmatrix} -720 \\ 2160 \\ 720 \\ 720 \end{pmatrix} = 1440\sqrt{3} \begin{pmatrix} -\frac{1}{6}\sqrt{3} \\ \frac{1}{2}\sqrt{3} \\ \frac{1}{6}\sqrt{3} \\ \frac{1}{6}\sqrt{3} \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 0 \\ 108 \\ -108 \end{pmatrix} = 108\sqrt{2} \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2}\sqrt{2} \\ -\frac{1}{2}\sqrt{2} \end{pmatrix}$$

$$\begin{pmatrix} -40 \\ -40 \\ 40 \\ 40 \end{pmatrix} = 80 \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\begin{pmatrix} -192 \\ 0 \\ -96 \\ -96 \end{pmatrix} = -96\sqrt{6} \begin{pmatrix} \frac{1}{3}\sqrt{6} \\ 0 \\ \frac{1}{6}\sqrt{6} \\ \frac{1}{6}\sqrt{6} \end{pmatrix}$$

Entonces, una matriz ortogonal P tal que $P^tQP = D_{\rho_1, \rho_2, \rho_3, \rho_4}$ está dada por:

$$P = \begin{pmatrix} -\frac{1}{6}\sqrt{3} & 0 & -\frac{1}{2} & \frac{1}{3}\sqrt{6} \\ \frac{1}{2}\sqrt{3} & 0 & -\frac{1}{2} & 0 \\ \frac{1}{6}\sqrt{3} & \frac{1}{2}\sqrt{2} & \frac{1}{2} & \frac{1}{6}\sqrt{6} \\ \frac{1}{6}\sqrt{3} & -\frac{1}{2}\sqrt{2} & \frac{1}{2} & \frac{1}{6}\sqrt{6} \end{pmatrix}$$

lo cual se puede verificar directamente:

$$QP = \begin{pmatrix} 2\sqrt{3} & 0 & -4 & 4\sqrt{6} \\ -6\sqrt{3} & 0 & -4 & 0 \\ -2\sqrt{3} & 3\sqrt{2} & 4 & 2\sqrt{6} \\ -2\sqrt{3} & -3\sqrt{2} & 4 & 2\sqrt{6} \end{pmatrix}$$

$$P^tQP = \begin{pmatrix} -12 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix}$$

▲

En general, para encontrar una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \dots, \alpha_n}$, en donde Q es una matriz simétrica de $n \times n$, primero se encuentran los valores propios de Q , es decir, las raíces del polinomio $p(z) = |Q - zI_n|$. Después se encuentran vectores propios correspondientes a los valores propios encontrados. Sabemos que si dos valores propios son distintos entonces dos vectores propios correspondientes, respectivamente, a esos valores propios, son ortogonales; pero si $p(z)$ tiene alguna raíz α de multiplicidad k , en donde $k > 1$, entonces k vectores propios correspondientes a α , no necesariamente son ortogonales, incluso cuando son linealmente independientes. En ese caso, se encuentran k vectores propios linealmente independientes correspondientes a α y se aplica a esos vectores el proceso de ortogonalización de Gram-Schmidt.

EJEMPLO 3.47. Sea $Q = \frac{1}{4} \begin{pmatrix} 5 & -1 & -1 & 1 \\ -1 & 5 & 1 & -1 \\ -1 & 1 & 5 & -1 \\ 1 & -1 & -1 & 5 \end{pmatrix}$

El determinante $|Q - zI_4|$ está dado por:

$$z^4 - 5z^3 + 9z^2 - 7z + 2 = (z - 2)(z - 1)^3$$

Para obtener un vector propio asociado al valor propio 1, se tiene que resolver el sistema de ecuaciones $Qx = x$ cuya matriz está dada por:

$$\begin{pmatrix} \frac{5}{4} - 1 & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{5}{4} - 1 & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{5}{4} - 1 & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{5}{4} - 1 \end{pmatrix}$$

la cual tiene la siguiente forma escalonada reducida:

$$\begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Por lo tanto, hay tres soluciones linealmente independientes, dadas por los vectores:

$$w_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, w_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, w_3 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Para obtener un vector propio asociado al valor propio 2, resolvamos el sistema de ecuaciones $Qx = 2x$ cuya matriz está dada por:

$$\begin{pmatrix} \frac{5}{4} - 2 & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{5}{4} - 2 & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{5}{4} - 2 & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{5}{4} - 2 \end{pmatrix}$$

la cual tiene la siguiente forma escalonada reducida:

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Por lo tanto, una solución está dada por el vector:

$$w_4 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

w_4 es ortogonal a w_1 , w_2 y w_3 ya que corresponden a distintos valores propios, pero w_1 , w_2 y w_3 no son ortogonales. Entonces, para obtener 3 vectores propios correspondientes al valor propio 1, aplicaremos el proceso de ortogonalización de Gram-Schmidt a w_1, w_2, w_3 .

Sean:

$$v_1 = w_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$v_2 = w_2 - \frac{1}{2}(w_2 \cdot v_1)v_1 = w_2 - \frac{1}{2}v_1 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}$$

$$v_3 = w_3 - \frac{1}{2}(w_3 \cdot v_1)v_1 - \frac{2}{3}(w_3 \cdot v_2)v_2 = w_3 + \frac{1}{2}v_1 + \frac{1}{3}v_2 = \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ 1 \end{pmatrix}$$

Los vectores v_1 , v_2 y v_3 son ortogonales y, por ser combinaciones lineales de w_1 , w_2 y w_3 , son también vectores propios correspondientes al valor propio 1. Así que una matriz ortogonal P tal que $P^tQP = D_{1,1,1,2}$ está dada por:

$$P = \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{6} & -\frac{1}{6}\sqrt{3} & \frac{1}{2} \\ \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{3} & -\frac{1}{2} \\ 0 & \frac{1}{3}\sqrt{6} & \frac{1}{6}\sqrt{3} & -\frac{1}{2} \\ 0 & 0 & \frac{1}{2}\sqrt{3} & \frac{1}{2} \end{pmatrix}$$

PROPOSICIÓN 3.48. Una matriz Q , de $n \times n$, es simétrica y definida positiva si y sólo si existen n números reales positivos $\alpha_1, \alpha_2, \dots, \alpha_n$ y una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \alpha_2, \dots, \alpha_n}$.

Demostración

Supongamos primero que existen n números reales positivos $\alpha_1, \dots, \alpha_n$ y una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \alpha_2, \dots, \alpha_n}$. Entonces $Q = PD_{\alpha_1, \dots, \alpha_n}P^t$, por lo tanto $Q^t = PD_{\alpha_1, \dots, \alpha_n}P^t = Q$, así que Q es simétrica. Además, $D_{\alpha_1, \dots, \alpha_n}$ es definida positiva, así que, si $x \neq 0$:

$$x^tQx = x^tPD_{\alpha_1, \dots, \alpha_n}P^tx = (P^tx)^tD_{\alpha_1, \dots, \alpha_n}(P^tx) > 0$$

Por lo tanto, Q es definida positiva.

Supongamos ahora que Q es simétrica y definida positiva.

Por la proposición 3.43, existen n números reales $\alpha_1, \dots, \alpha_n$ y una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \dots, \alpha_n}$.

Para cada $k \in \{1, \dots, n\}$, sea x_k el vector que forma la k -ésima columna de P y w_k el vector de \mathbb{R}^n cuya k -ésima coordenada es 1 y todas las demás son cero. Entonces, como P es invertible, $Pw_k \neq 0$ y como Q es definida positiva, $(Pw_k)^tQ(Pw_k) > 0$. Por lo tanto:

$$\alpha_k = w_k^t\alpha_k w_k = w_k^tD_{\alpha_1, \dots, \alpha_n}w_k = w_k^tP^tQPw_k = (Pw_k)^tQ(Pw_k) > 0$$

■

COROLARIO 3.49. Una matriz simétrica Q es definida positiva si y sólo si sus valores propios son positivos.

EJEMPLO 3.50. Sea $Q = \begin{pmatrix} 11 & 9 & 0 & 0 \\ 9 & 11 & 0 & 0 \\ 0 & 0 & 9 & -2 \\ 0 & 0 & -2 & 6 \end{pmatrix}$

El determinante $|Q - zI_n|$ está dado por:

$$p(z) = z^4 - 37z^3 + 420z^2 - 1700z + 2000$$

Las raíces de este polinomio son $\rho_1 = 2$, $\rho_2 = 5$, $\rho_3 = 10$, y $\rho_4 = 20$, así que Q es definida positiva.

Sea ρ una de las raíces de p . Para obtener un vector propio asociado a ρ , se tiene que resolver el sistema de ecuaciones $Qx = \rho x$ cuya matriz está dada por:

$$\begin{pmatrix} 11 - \rho & 9 & 0 & 0 \\ 9 & 11 - \rho & 0 & 0 \\ 0 & 0 & 9 - \rho & -2 \\ 0 & 0 & -2 & 6 - \rho \end{pmatrix}$$

la cual se puede llevar a su forma escalonada reducida de la siguiente manera:

$$\begin{aligned} &\rightarrow \begin{pmatrix} 9 & 11 - \rho & 0 & 0 \\ 11 - \rho & 9 & 0 & 0 \\ 0 & 0 & 9 - \rho & -2 \\ 0 & 0 & -2 & 6 - \rho \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 9 & 11 - \rho & 0 & 0 \\ 0 & -\frac{(\rho-2)(\rho-20)}{9} & 0 & 0 \\ 0 & 0 & 9 - \rho & -2 \\ 0 & 0 & -2 & 6 - \rho \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 9 & 11 - \rho & 0 & 0 \\ 0 & (\rho-2)(\rho-20) & 0 & 0 \\ 0 & 0 & -2 & 6 - \rho \\ 0 & 0 & 9 - \rho & -2 \end{pmatrix} \\ &\rightarrow \begin{pmatrix} 1 & \frac{-\rho+11}{9} & 0 & 0 \\ 0 & (\rho-2)(\rho-20) & 0 & 0 \\ 0 & 0 & 1 & \frac{\rho-6}{2} \\ 0 & 0 & 0 & (\rho-5)(\rho-10) \end{pmatrix} \end{aligned}$$

Así que, para $\rho = 2$ o $\rho = 20$, se obtiene la matriz:

$$\begin{pmatrix} 1 & \frac{-\rho+11}{9} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \frac{-\rho+11}{9} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Por lo tanto, en este caso, una solución está dada por el vector:

$$v = \begin{pmatrix} \frac{\rho-11}{9} \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Es decir, dos vectores propios correspondientes a $\rho = 2$ y $\rho = 20$ están dados, respectivamente, por:

$$\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \text{ y } \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

para $\rho = 5$ o $\rho = 10$, se obtiene la matriz:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{\rho-6}{2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Por lo tanto, en este caso, una solución está dada por el vector:

$$w = \begin{pmatrix} 0 \\ 0 \\ -\frac{\rho-6}{2} \\ 1 \end{pmatrix}$$

Es decir, dos vectores propios correspondientes a $\rho = 5$ y $\rho = 10$ están dados, respectivamente, por:

$$\begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ 1 \end{pmatrix} \text{ y } \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}$$

Así que una matriz ortogonal P tal que $P^tQP = D_{\rho_1, \rho_2, \rho_3, \rho_4}$ está dada por:

$$P = \begin{pmatrix} -\frac{1}{2}\sqrt{2} & 0 & 0 & \frac{1}{2}\sqrt{2} \\ \frac{1}{2}\sqrt{2} & 0 & 0 & \frac{1}{2}\sqrt{2} \\ 0 & \frac{1}{5}\sqrt{5} & -\frac{2}{5}\sqrt{5} & 0 \\ 0 & \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \end{pmatrix}$$

▲

Las matrices simétricas y definidas positivas tienen propiedades similares a las de los números reales positivos. Un ejemplo de ello es la siguiente proposición, la cual establece una de sus principales propiedades.

PROPOSICIÓN 3.51. *Sea Q una matriz simétrica y definida positiva, entonces existe una matriz B , invertible, simétrica y definida positiva tal que $Q = B^2$.*

Demostración

Por la proposición 3.48, si Q es de $n \times n$, existen n números reales positivos $\alpha_1, \dots, \alpha_n$ y una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \dots, \alpha_n}$.

Sea D la matriz diagonal formada por las raíces positivas de $\alpha_1, \dots, \alpha_n$, es decir, $D = D_{\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n}}$, y definamos $B = PDP^t$. Entonces:

$$B^2 = PDP^tPDP^t = PD^2P^t = PD_{\alpha_1, \dots, \alpha_n}P^t = Q$$

■

EJEMPLO 3.52. Sea $Q = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{10}{9} & \frac{4}{9} & \frac{3}{9} \\ \frac{1}{3} & \frac{4}{9} & \frac{11}{9} & \frac{9}{9} \\ \frac{1}{3} & \frac{9}{9} & \frac{9}{9} & \frac{9}{3} \end{pmatrix}$

El determinante $|Q - zI_4|$ está dado por:

$$p(z) = z^4 - \frac{14}{3}z^3 + \frac{575}{81}z^2 - \frac{3247}{729}z + 1.$$

Las raíces ρ_1, ρ_2, ρ_3 y ρ_4 de p pueden estimarse mediante algún método numérico, obteniéndose:

$$\rho_1 \approx 0.69894458127259582741$$

$$\rho_2 \approx 0.71972257623022239805$$

$$\rho_3 \approx 0.81809269002976566574$$

$$\rho_4 \approx 2.4299068191340827755$$

Así que Q es definida positiva.

Sea ρ una de las raíces de p . Para obtener un vector propio asociado a ρ , se tiene que resolver el sistema de ecuaciones $Qx = \rho x$ cuya matriz está dada por:

$$\begin{pmatrix} 1 - \rho & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{10}{9} - \rho & \frac{4}{9} & \frac{4}{9} \\ \frac{1}{3} & \frac{4}{9} & \frac{11}{9} - \rho & \frac{4}{9} \\ \frac{1}{3} & \frac{4}{9} & \frac{4}{9} & \frac{4}{3} - \rho \end{pmatrix}$$

la cual, al reducirla, siguiendo el método de Gauss-Jordan, puede llevarse a la forma siguiente:

$$\rightarrow \begin{pmatrix} \frac{1}{3} & \frac{10}{9} - \rho & \frac{4}{9} & \frac{4}{9} \\ 0 & -\frac{2}{3} + \rho & \frac{7}{9} - \rho & \frac{4}{9} \\ 0 & 0 & -\frac{2}{3} + \rho & \frac{7}{9} - \rho \\ 0 & 0 & 0 & -\frac{1}{27} \frac{729\rho^4 - 3402\rho^3 + 5175\rho^2 - 3247\rho + 729}{(2-3\rho)^2} \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} \frac{1}{3} & \frac{10}{9} - \rho & \frac{4}{9} & \frac{4}{9} \\ 0 & -\frac{2}{3} + \rho & \frac{7}{9} - \rho & \frac{4}{9} \\ 0 & 0 & -\frac{2}{3} + \rho & \frac{7}{9} - \rho \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{27} \frac{729\rho^3 - 2673\rho^2 + 2745\rho - 862}{(3\rho - 2)^2} \\ 0 & 1 & 0 & -\frac{1}{9} \frac{81\rho^2 - 135\rho + 55}{(3\rho - 2)^2} \\ 0 & 0 & 1 & -\frac{1}{3} \frac{9\rho - 7}{3\rho - 2} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Por lo tanto, una solución está dada por el vector:

$$w = \begin{pmatrix} \frac{1}{27} \frac{729\rho^3 - 2673\rho^2 + 2745\rho - 862}{(3\rho - 2)^2} \\ \frac{1}{9} \frac{81\rho^2 - 135\rho + 55}{(3\rho - 2)^2} \\ \frac{1}{3} \frac{9\rho - 7}{3\rho - 2} \\ 1 \end{pmatrix}$$

Así que, una matriz ortogonal P tal que $P^t Q P = D_{\rho_1, \rho_2, \rho_3, \rho_4}$ está dada por:

$$P \approx \begin{pmatrix} -.3113471995 & .5625057937 & -.6688159463 & .3732765507 \\ .6566290227 & -.4280070195 & -.401000884 & .474179933 \\ -.6357260432 & -.5221760344 & .161054579 & .5452027235 \\ .2602953051 & .4772100297 & .604933834 & .5818693995 \end{pmatrix}$$

Por lo tanto, una matriz B , invertible, simétrica y definida positiva tal que $Q = B^2$ está dada por:

$$B = P\sqrt{D}P^t \approx \begin{pmatrix} .9712632179 & .1433231007 & .1360982808 & .1326028212 \\ .1433231008 & 1.011812765 & .1851937348 & .1802992213 \\ .1360982808 & .1851937348 & 1.056013705 & .2328901348 \\ .1326028211 & .1802992212 & .2328901349 & 1.108604619 \end{pmatrix}$$

3.3. Distribución normal multivariada

Si la pareja de variables aleatorias X, Y tiene distribución normal bivariada, existen dos variables aleatorias independientes, U, V , con distribución normal estándar, tales que $X = aU + bV + \mu_1$ y $Y = cU + dV + \mu_2$, en donde a, b, c, d, μ_1 y μ_2 son constantes tales que $ad - bc \neq 0$. Esta propiedad puede expresarse diciendo que existe una matriz invertible $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ y un vector $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ tales que:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A \begin{pmatrix} U \\ V \end{pmatrix} + \mu$$

Esta forma de expresar la propiedad que caracteriza a una distribución normal bivariada permite extender la definición al caso de tres o más variables aleatorias.

DEFINICIÓN 3.53 (Distribución normal multivariada). Se dice que la familia de variables aleatorias X_1, \dots, X_n tiene distribución normal multivariada si existen n variables aleatorias independientes U_1, \dots, U_n , todas con distribución normal estándar, una matriz de $n \times n$ invertible A y un vector n -dimensional μ tales que $X = AU + \mu$,

$$\text{en donde } X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \text{ y } U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}.$$

PROPOSICIÓN 3.54. Supongamos que la familia de variables aleatorias X_1, \dots, X_n tiene distribución normal multivariada y sean A, U y μ tales que $X = AU + \mu$, de acuerdo con la definición 3.53, entonces μ y $C = AA^t$ son el vector de esperanzas y la matriz de covarianzas, respectivamente, de X_1, \dots, X_n .

Demostración

Sean $A = (a_{ij})$ y $C = (c_{ij})$, entonces $X_i = \sum_{k=1}^n a_{ik}U_k + \mu_i$, así que $E[X_i] = \mu_i$ y

$$E[(X_i - \mu_i)(X_j - \mu_j)] = E[(\sum_{k=1}^n a_{ik}U_k)(\sum_{k=1}^n a_{jk}U_k)] = \sum_{k=1}^n a_{ik}a_{jk} = c_{ij}. \quad \blacksquare$$

PROPOSICIÓN 3.55. *Supongamos que la familia de variables aleatorias X_1, \dots, X_n tiene distribución normal multivariada y sean A , U y μ tales que $X = AU + \mu$, de acuerdo con la definición 3.53, entonces una función de densidad conjunta de X_1, \dots, X_n está dada por:*

$$f_{X_1, X_2, \dots, X_n}(x) = \frac{\sqrt{|C^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} C^{-1}(x - \mu) \cdot (x - \mu) \right\}$$

en donde $C = AA^t$.

Demostración

Para encontrar una función de densidad conjunta de X_1, \dots, X_n , consideremos la transformación $x = Au + \mu$, la cual tiene como inversa a $u = A^{-1}(x - \mu)$, cuyo jacobiano está dado por $|A^{-1}|$, de manera que:

$$\begin{aligned} f_{X_1, \dots, X_n}(x) &= |A^{-1}| f_{U_1, \dots, U_n}(A^{-1}(x - \mu)) \\ &= \frac{\sqrt{|A^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} A^{-1}(x - \mu) \cdot A^{-1}(x - \mu) \right\} \\ &= \frac{\sqrt{|A^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} [A^{-1}(x - \mu)]^t A^{-1}(x - \mu) \right\} \\ &= \frac{\sqrt{|A^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} (x - \mu)^t (A^{-1})^t A^{-1}(x - \mu) \right\} \\ &= \frac{\sqrt{|A^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} (x - \mu)^t (A^t)^{-1} A^{-1}(x - \mu) \right\} \\ &= \frac{\sqrt{|A^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} (x - \mu)^t (AA^t)^{-1} (x - \mu) \right\} \\ &= \frac{\sqrt{|C^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} C^{-1}(x - \mu) \cdot (x - \mu) \right\} \end{aligned}$$

■

De acuerdo con las proposiciones 3.55 y 3.29, si la familia de variables aleatorias X_1, \dots, X_n tiene distribución normal multivariada entonces una función de densidad conjunta, f_{X_1, \dots, X_n} , de X_1, \dots, X_n está dada por:

$$f_{X_1, \dots, X_n}(x) = \frac{\sqrt{|C^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} C^{-1}(x - \mu) \cdot (x - \mu) \right\}$$

en donde C es una matriz simétrica y definida positiva. La siguiente proposición muestra que el inverso de este resultado es también válido.

PROPOSICIÓN 3.56. *Sea X un vector aleatorio con función de densidad conjunta f_X dada por:*

$$f_X(x) = K \exp \left\{ -\frac{1}{2} C^{-1}(x - \mu) \cdot (x - \mu) \right\}$$

en donde C es una matriz simétrica y definida positiva y K y μ son constantes, entonces X tiene distribución normal multivariada.

Demostración

Sea A una matriz invertible tal que $C = AA^t$ y consideremos la transformación $x = Au + \mu$, entonces:

$$\begin{aligned} (C^{-1}(Au)) \cdot (Au) &= ((A^t)^{-1}A^{-1}(Au)) \cdot (Au) = ((A^t)^{-1}(u)) \cdot (Au) \\ &= (Au)^t ((A^t)^{-1}(u)) = u^t A^t (A^t)^{-1}u = u^t u = u \cdot u \end{aligned}$$

Así que, definiendo $U = A^{-1}(X - \mu)$, se tiene:

$$f_U(u) = |A| f_X(\mu + Au) = |A| K \exp \left\{ -\frac{1}{2} (C^{-1}(Au)) \cdot (Au) \right\} = |A| K \exp \left\{ -\frac{1}{2} u \cdot u \right\}$$

Por lo tanto, el vector aleatorio U está formado por variables aleatorias independientes, todas con distribución normal estándar. Además, $X = AU + \mu$ y la matriz A es invertible, así que X tiene distribución normal multivariada. ■

Sea $f : \mathbb{R}^n \mapsto \mathbb{R}$ una función de densidad de la forma:

$$f(x_1, \dots, x_n) = C \exp \left\{ -\frac{1}{2} \left(\sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} x_i x_j + \sum_{i=1}^n a_i x_i \right) \right\}$$

en donde a_1, \dots, a_n y C son constantes y la forma cuadrática $F(x_1, \dots, x_n) = \sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} x_i x_j$ es definida positiva.

La matriz asociada a la forma cuadrática F está dada por:

$$Q = \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

Para $i, j \in \{1, \dots, n\}$, con $i > j$, definamos $a_{ij} = a_{ji}$. Entonces, para μ_1, \dots, μ_n números reales cualesquiera, se tiene:

$$F(x_1 - \mu_1, \dots, x_n - \mu_n) = \sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} (x_i - \mu_i)(x_j - \mu_j)$$

$$\begin{aligned}
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i x_j \\
&- \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_j x_i + \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j \\
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{j=1}^n \sum_{i=1}^j a_{ij} \mu_i x_j - \sum_{i=1}^n \sum_{j=1}^n a_{ij} \mu_j x_i \\
&+ \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j \\
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{i=1}^n \sum_{j=1}^i a_{ji} \mu_j x_i - \sum_{i=1}^n \sum_{j=1}^n a_{ij} \mu_j x_i \\
&+ \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j \\
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{i=1}^n \left(\sum_{j=1}^i a_{ji} \mu_j + \sum_{j=i}^n a_{ij} \mu_j \right) x_i + \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j \\
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{i=1}^n \left(\sum_{j=1}^i a_{ij} \mu_j + \sum_{j=i}^n a_{ij} \mu_j \right) x_i + \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j \\
&= \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} x_i x_j - \sum_{i=1}^n \left(a_{ii} \mu_i + \sum_{j=1}^n a_{ij} \mu_j \right) x_i + \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} \mu_i \mu_j
\end{aligned}$$

Consideremos entonces el sistema de ecuaciones en μ_1, \dots, μ_n :

$$a_{ii} \mu_i + \sum_{j=1}^n a_{ij} \mu_j = -a_i$$

El determinante de este sistema está dado por:

$$\begin{vmatrix} 2a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & 2a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & 2a_{nn} \end{vmatrix} = 2^n \begin{vmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{vmatrix} = 2^n |Q| \neq 0$$

Así que el sistema tiene una solución única.

Por lo tanto, f puede escribirse en la forma siguiente:

$$f(x_1, \dots, x_n) = K \exp \left\{ -\frac{1}{2} \sum_{\{i,j \in \{1, \dots, n\} : i \leq j\}} a_{ij} (x_i - \mu_i)(x_j - \mu_j) \right\}$$

en donde K es una constante.

Así que f es función de densidad de una distribución normal multivariada.

Se tiene entonces el siguiente resultado:

PROPOSICIÓN 3.57. *Sea $f : \mathbb{R}^2 \mapsto \mathbb{R}$ una función de densidad de la forma:*

$$f(x_1, \dots, x_n) = C \exp \left\{ -\frac{1}{2} \left(\sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} x_i x_j + \sum_{i=1}^n a_i x_i \right) \right\}$$

en donde a_1, \dots, a_n y C son constantes y la forma cuadrática:

$$F(x_1, \dots, x_n) = \sum_{\{i,j \in \{1, \dots, n\}: i \leq j\}} a_{ij} x_i x_j$$

es definida positiva. Entonces f es función de densidad de una distribución normal multivariada.

EJEMPLO 3.58. Sea (X_1, X_2, X_3, X_4) un vector aleatorio con función de densidad conjunta f dada por

$$\begin{aligned} f(x_1, x_2, x_3, x_4) \\ = C \exp \left\{ -\frac{1}{2} (x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1 x_2 - x_2 x_3 - x_3 x_4 + 2x_1 - 3x_2 - x_3 + x_4) \right\} \end{aligned}$$

en donde C es una constante.

Completando cuadrados, se tiene:

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1 x_2 - x_2 x_3 - x_3 x_4 \\ = \left(x_1 - \frac{1}{2}x_2\right)^2 + \frac{3}{4} \left(x_2 - \frac{2}{3}x_3\right)^2 + \frac{2}{3} \left(x_3 - \frac{3}{4}x_4\right)^2 + \frac{5}{8}x_4^2 \end{aligned}$$

Así que la forma cuadrática:

$$F(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1 x_2 - x_2 x_3 - x_3 x_4$$

es definida positiva.

Sean $\mu_1, \mu_2, \mu_3, \mu_4$ las esperanzas de X_1, X_2, X_3, X_4 , respectivamente, entonces,

$$\begin{aligned} F(x_1 - \mu_1, x_2 - \mu_2, x_3 - \mu_3, x_4 - \mu_4) \\ = (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 + (x_3 - \mu_3)^2 + (x_4 - \mu_4)^2 \\ - (x_1 - \mu_1)(x_2 - \mu_2) - (x_2 - \mu_2)(x_3 - \mu_3) - (x_3 - \mu_3)(x_4 - \mu_4) \\ = x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1 x_2 - x_2 x_3 - x_3 x_4 \\ + (-2\mu_1 + \mu_2)x_1 + (\mu_1 - 2\mu_2 + \mu_3)x_2 + (\mu_2 - 2\mu_3 + \mu_4)x_3 + (\mu_3 - 2\mu_4)x_4 \\ + \mu_1^2 + \mu_2^2 + \mu_3^2 + \mu_4^2 - \mu_1 \mu_2 - \mu_2 \mu_3 - \mu_3 \mu_4 \end{aligned}$$

Así que:

$$-2\mu_1 + \mu_2 = 2$$

$$\mu_1 - 2\mu_2 + \mu_3 = -3$$

$$\mu_2 - 2\mu_3 + \mu_4 = -1$$

$$\mu_3 - 2\mu_4 = 1$$

Por lo tanto:

$$\mu_1 = \frac{2}{5}, \mu_2 = \frac{14}{5}, \mu_3 = \frac{11}{5}, \mu_4 = \frac{3}{5}$$

Así que:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) &= K \exp \left\{ -\frac{1}{2} F(x_1 - \frac{2}{5}, x_2 - \frac{14}{5}, x_3 - \frac{11}{5}, x_4 - \frac{3}{5}) \right\} \\ &= K \exp \left\{ -\frac{1}{2} \left[(x_1 - \frac{2}{5})^2 + (x_2 - \frac{14}{5})^2 + (x_3 - \frac{11}{5})^2 + (x_4 - \frac{3}{5})^2 \right. \right. \\ &\quad \left. \left. - (x_1 - \frac{2}{5})(x_2 - \frac{14}{5}) - (x_2 - \frac{14}{5})(x_3 - \frac{11}{5}) - (x_3 - \frac{11}{5})(x_4 - \frac{3}{5}) \right] \right\} \end{aligned}$$

en donde K es una constante.

La matriz Q asociada a la forma cuadrática F está dada por:

$$Q = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

Así que la matriz de covarianzas de X_1, X_2, X_3, X_4 está dada por:

$$C = Q^{-1} = \begin{pmatrix} \frac{8}{5} & \frac{6}{5} & \frac{4}{5} & \frac{2}{5} \\ \frac{6}{5} & \frac{12}{5} & \frac{8}{5} & \frac{2}{5} \\ \frac{4}{5} & \frac{8}{5} & \frac{12}{5} & \frac{4}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{4}{5} & \frac{4}{5} \end{pmatrix} = \frac{2}{5} \begin{pmatrix} 4 & 3 & 2 & 1 \\ 3 & 6 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

Además:

$$K = \frac{\sqrt{|C^{-1}|}}{(\sqrt{2\pi})^4} = \frac{1}{\pi^2 \sqrt{5}}$$

Por lo tanto, si μ es el vector con coordenadas $\frac{2}{5}, \frac{14}{5}, \frac{11}{5}, \frac{3}{5}$, se tiene:

$$\begin{aligned} f(x) &= \frac{1}{\pi^2 \sqrt{5}} \exp \left\{ -\frac{1}{2} (x - \mu)^t C^{-1} (x - \mu) \right\} \\ &= \frac{1}{\pi^2 \sqrt{5}} \exp \left\{ -\frac{1}{2} \left[(x_1 - \frac{2}{5})^2 + (x_2 - \frac{14}{5})^2 + (x_3 - \frac{11}{5})^2 + (x_4 - \frac{3}{5})^2 \right. \right. \\ &\quad \left. \left. - (x_1 - \frac{2}{5})(x_2 - \frac{14}{5}) - (x_2 - \frac{14}{5})(x_3 - \frac{11}{5}) - (x_3 - \frac{11}{5})(x_4 - \frac{3}{5}) \right] \right\} \end{aligned}$$

$$-(x_1 - \frac{2}{5})(x_2 - \frac{14}{5}) - (x_2 - \frac{14}{5})(x_3 - \frac{11}{5}) - (x_3 - \frac{11}{5})(x_4 - \frac{3}{5}) \Big\}$$

Para verificar los cálculos anteriores, evaluemos el producto $(x - \mu)^t C^{-1} (x - \mu)$:

$$\begin{aligned} & (x - \mu)^t C^{-1} (x - \mu) \\ &= \left(x_1 - \frac{2}{5} \quad x_2 - \frac{14}{5} \quad x_3 - \frac{11}{5} \quad x_4 - \frac{3}{5} \right) \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x_1 - \frac{2}{5} \\ x_2 - \frac{14}{5} \\ x_3 - \frac{11}{5} \\ x_4 - \frac{3}{5} \end{pmatrix} \\ &= x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_1x_2 - x_2x_3 - x_3x_4 + 2x_1 - 3x_2 - x_3 + x_4 + \frac{23}{5} \end{aligned}$$

PROPOSICIÓN 3.59. Sea X un vector aleatorio n -dimensional con distribución normal multivariada con vector de esperanzas μ y matriz de covarianzas C . Sea ν un vector n -dimensional y A una matriz de $n \times n$ invertible. Entonces $Y = AX + \nu$ tiene distribución normal multivariada con vector de esperanzas $A\mu + \nu$ y matriz de covarianzas ACA^t .

Demostración

Sabemos que $X = BZ + \mu$, en donde B es una matriz invertible tal que $C = BB^t$ y Z es un vector aleatorio n -dimensional formado por variables aleatorias independientes con distribución normal estándar. Por lo tanto:

$$Y = A(BZ + \mu) + \nu = (AB)Z + (A\mu + \nu)$$

De manera que Y tiene distribución normal multivariada con matriz de covarianzas $C_Y = (AB)(AB)^t = ABB^tA^t = ACA^t$ y vector de esperanzas $\mu_Y = A\mu + \nu$. ■

Una propiedad importante de una distribución normal multivariada consiste en que basta con que sea nula la covarianza entre cada pareja de variables aleatorias de la familia para asegurar que son independientes. Esto se prueba a continuación.

PROPOSICIÓN 3.60. Sea $X = (X_1, \dots, X_n)$ un vector aleatorio con distribución normal multivariada con vector de esperanzas μ y tal que su matriz de covarianzas C es diagonal. Entonces las variables aleatorias X_1, \dots, X_n son independientes.

Demostración

Una función de densidad conjunta de X_1, \dots, X_n está dada por:

$$f_{X_1, \dots, X_n}(\bar{x}) = \frac{\sqrt{|C^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} C^{-1} (\bar{x} - \bar{\mu}) \cdot (\bar{x} - \bar{\mu}) \right\}$$

en donde $\bar{\mu}$ es el vector de esperanzas de X_1, X_2, \dots, X_n .

Como C es diagonal, tiene la forma:

$$C = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

en donde $\sigma_1^2, \dots, \sigma_n^2$ son las varianzas de X_1, \dots, X_n , respectivamente. Por lo tanto:

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n) &= \frac{1}{(\sqrt{2\pi})^n \sigma_1 \cdots \sigma_n} \exp \left\{ -\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2 - \cdots - \frac{1}{2\sigma_n^2}(x_n - \mu_n)^2 \right\} \\ &= \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2 \right\} \cdots \frac{1}{\sigma_n \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_n^2}(x_n - \mu_n)^2 \right\} \\ &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \end{aligned}$$

Así que las variables aleatorias X_1, \dots, X_n son independientes. ■

COROLARIO 3.61. Sean U_1, \dots, U_n n variables aleatorias independientes, todas con distribución normal estándar, P una matriz de $n \times n$ ortogonal y V_1, \dots, V_n las variables aleatorias definidas mediante la relación:

$$\begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} = P \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

Entonces V_1, \dots, V_n son independientes y todas tienen distribución normal estándar.

Demostración

La matriz de covarianzas de la familia V_1, \dots, V_n está dada por $C = AA^t$, así que, como A es ortogonal, C es la identidad. El resultado se sigue entonces de la proposición 3.60. ■

De acuerdo con la definición 3.53, si el vector aleatorio n -dimensional X tiene distribución normal multivariada, existe un vector aleatorio n -dimensional U formado por variables aleatorias independientes, todas con distribución normal estándar, una matriz de $n \times n$ invertible A y un vector n -dimensional μ tales que $X = AU + \mu$.

La pareja U, A en esta representación, no es única. En efecto, si P es una matriz ortogonal, entonces, de acuerdo con la proposición 3.61, el vector aleatorio $U' = PU$ también está formado por variables aleatorias independientes con distribución normal estándar y, definiendo $A' = AP^t$, A' es invertible y se tiene: $A'U' = AP^tPU = AU$.

Dado un vector aleatorio X con distribución normal multivariada, con matriz de covarianzas C y vector de esperanzas μ , para expresar X en la forma $X = AU + \mu$, en donde A es una matriz invertible y U un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar, se requiere encontrar una matriz invertible A tal que $C = AA^t$ y entonces se tiene $X = AU + \mu$, en donde U es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar. En efecto, Si A es una matriz de $n \times n$ invertible tal que $C = AA^t$, entonces $U = A^{-1}(X - \mu)$ tiene distribución normal multivariada con matriz de covarianzas:

$$C_U = A^{-1}C(A^{-1})^t = A^{-1}C(A^t)^{-1} = A^{-1}AA^t(A^t)^{-1} = I_n$$

y vector de esperanzas $\mu_U = A^{-1}\mu - A^{-1}\mu = 0$

Así que U es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar y $X = AU + \mu$.

Ahora bien, para encontrar una matriz invertible A tal que $C = AA^t$ se pueden seguir diferentes procedimientos. Uno de ellos consiste en encontrar una matriz ortogonal P tal que $P^tCP = D$, en donde D es una matriz diagonal cuyos elementos sobre la diagonal son positivos. De acuerdo con la proposición 3.48, tales matrices existen. Entonces tanto $A = P\sqrt{D}$ como $A' = P\sqrt{D}P^t$ satisfacen la propiedad requerida. En efecto, se tiene:

$$AA^t = \left(P\sqrt{D}\right) \left(P\sqrt{D}\right)^t = P\sqrt{D}\sqrt{D}P^t = PDP^t = C$$

$$A'(A')^t = \left(P\sqrt{D}P^t\right) \left(P\sqrt{D}P^t\right)^t = P\sqrt{D}P^tP\sqrt{D}P^t = P\sqrt{D}\sqrt{D}P^t = PDP^t = C$$

Este método para expresar X en la forma $X = AU + \mu$ puede resultar muy laborioso pues para encontrar P se requiere encontrar los valores y vectores propios de C , lo cual no siempre resulta un proceso simple.

En general, el método más simple para encontrar una matriz invertible A tal que $C = AA^t$ consiste en aplicar a la forma cuadrática definida por C^{-1} el método de Lagrange, de completación de cuadrados, el cual, de acuerdo con la proposición 3.35, nos permite encontrar una matriz triangular superior invertible B tal que $B^tB = C^{-1}$. Definiendo entonces $A = B^{-1}$, se tiene:

$$AA^t = (B^{-1})(B^{-1})^t = (B^t B)^{-1} = C$$

EJEMPLO 3.62. Sea X un vector aleatorio con distribución normal multivariada, con vector de esperanzas 0 y matriz de covarianzas C , en donde:

$$C = \begin{pmatrix} 1 & \frac{1}{3} & 0 \\ \frac{1}{3} & 1 & \frac{2}{9} \\ 0 & \frac{2}{9} & 1 \end{pmatrix}$$

Se tiene:

$$C^{-1} = \begin{pmatrix} \frac{77}{68} & -\frac{27}{68} & \frac{3}{34} \\ -\frac{27}{68} & \frac{81}{68} & -\frac{9}{34} \\ \frac{3}{34} & -\frac{9}{34} & \frac{18}{17} \end{pmatrix}$$

La forma cuadrática definida por C^{-1} está dada por:

$$\begin{aligned} F(x_1, x_2, x_3) &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} \frac{77}{68} & -\frac{27}{68} & \frac{3}{34} \\ -\frac{27}{68} & \frac{81}{68} & -\frac{9}{34} \\ \frac{3}{34} & -\frac{9}{34} & \frac{18}{17} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= \frac{77}{68}x_1^2 - \frac{27}{34}x_1x_2 + \frac{3}{17}x_1x_3 + \frac{81}{68}x_2^2 - \frac{9}{17}x_2x_3 + \frac{18}{17}x_3^2 \\ &= \frac{77}{68} \left(x_1 - \frac{27}{77}x_2 + \frac{6}{77}x_3 \right)^2 - \frac{77}{68} \left(-\frac{27}{77}x_2 + \frac{6}{77}x_3 \right)^2 + \frac{81}{68}x_2^2 - \frac{9}{17}x_2x_3 + \frac{18}{17}x_3^2 \\ &= \frac{77}{68} \left(x_1 - \frac{27}{77}x_2 + \frac{6}{77}x_3 \right)^2 + \frac{81}{77} \left(x_2 - \frac{2}{9}x_3 \right)^2 + x_3^2 \end{aligned}$$

Así que si definimos:

$$A = \begin{pmatrix} \frac{\sqrt{77}}{2\sqrt{17}} & -\frac{27}{2\sqrt{1309}} & \frac{3}{\sqrt{1309}} \\ 0 & \frac{9}{\sqrt{77}} & -\frac{2}{\sqrt{77}} \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2\sqrt{17}}{\sqrt{77}} & \frac{3}{\sqrt{77}} & 0 \\ 0 & \frac{\sqrt{77}}{9} & \frac{2}{9} \\ 0 & 0 & 1 \end{pmatrix}$$

entonces $U = A^{-1}X$ es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar y $X = AU$.

EJEMPLO 3.63. Sea X un vector aleatorio con distribución normal multivariada, con vector de esperanzas 0 y matriz de covarianzas C , en donde:

$$C = \begin{pmatrix} \frac{73}{8} & \frac{5}{33} & -\frac{13}{12} & -\frac{10}{3} & -\frac{29}{12} \\ \frac{5}{33} & \frac{32}{8} & \frac{11}{8} & -\frac{4}{4} & -\frac{12}{8} \\ -\frac{13}{12} & \frac{11}{8} & 3 & -1 & 0 \\ -\frac{10}{3} & -\frac{3}{8} & -1 & 2 & 1 \\ -\frac{29}{12} & -\frac{4}{8} & 0 & 1 & 1 \end{pmatrix}$$

Se tiene:

$$C^{-1} = \begin{pmatrix} 3 & -2 & 4 & 6 & \frac{1}{2} \\ -2 & \frac{16}{3} & -\frac{14}{3} & -\frac{9}{2} & \frac{5}{3} \\ 4 & -\frac{14}{3} & \frac{41}{3} & \frac{35}{2} & -\frac{5}{3} \\ 6 & -\frac{9}{2} & \frac{6}{35} & \frac{4}{217} & -\frac{6}{3} \\ \frac{1}{2} & \frac{5}{3} & -\frac{5}{6} & -\frac{4}{3} & \frac{43}{12} \end{pmatrix}$$

La forma cuadrática definida por C^{-1} está dada por:

$$\begin{aligned} F(x_1, x_2, x_3, x_4, x_5) &= \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{pmatrix} \begin{pmatrix} 3 & -2 & 4 & 6 & \frac{1}{2} \\ -2 & \frac{16}{3} & -\frac{14}{3} & -\frac{9}{2} & \frac{5}{3} \\ 4 & -\frac{14}{3} & \frac{41}{3} & \frac{35}{2} & -\frac{5}{3} \\ 6 & -\frac{9}{2} & \frac{6}{35} & \frac{4}{217} & -\frac{6}{3} \\ \frac{1}{2} & \frac{5}{3} & -\frac{5}{6} & -\frac{4}{3} & \frac{43}{12} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \\ &= 3x_1^2 + \frac{16}{3}x_2^2 + \frac{41}{6}x_3^2 + \frac{217}{16}x_4^2 + \frac{43}{12}x_5^2 - 4x_1x_2 + 8x_1x_3 + 12x_1x_4 + x_1x_5 \\ &\quad - \frac{28}{3}x_2x_3 - 9x_2x_4 + \frac{10}{3}x_2x_5 + \frac{35}{2}x_3x_4 - \frac{5}{3}x_3x_5 - \frac{3}{2}x_4x_5 \\ &= 3 \left(x_1 - \frac{2}{3}x_2 + \frac{4}{3}x_3 + 2x_4 + \frac{1}{6}x_5 \right)^2 + 4 \left(x_2 - \frac{1}{2}x_3 - \frac{1}{8}x_4 + \frac{1}{2}x_5 \right)^2 \\ &\quad + \frac{1}{2} \left(x_3 + x_4 - x_5 \right)^2 + (x_4 - x_5)^2 + x_5^2 \end{aligned}$$

Así que si definimos:

$$\begin{aligned} A &= \begin{pmatrix} \sqrt{3} & -\frac{2}{3}\sqrt{3} & \frac{4}{3}\sqrt{3} & 2\sqrt{3} & \frac{1}{6}\sqrt{3} \\ 0 & 2 & -1 & -\frac{1}{4} & 1 \\ 0 & 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{1}{3}\sqrt{3} & \frac{1}{3} & -\sqrt{2} & -\frac{11}{12} & -\frac{29}{12} \\ 0 & \frac{1}{2} & \frac{1}{2}\sqrt{2} & -\frac{3}{8} & -\frac{3}{8} \\ 0 & 0 & \sqrt{2} & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

entonces $U = A^{-1}X$ es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar y $X = AU$.

▲

Queremos demostrar ahora que si X_1, \dots, X_n es una familia de variables aleatorias con distribución normal multivariada y X_{j_1}, \dots, X_{j_k} es una subfamilia de esa colección, entonces X_{j_1}, \dots, X_{j_k} también tiene distribución normal multivariada.

Para eso, obsérvese primero que si X_1, \dots, X_n tiene distribución normal multivariada entonces, de acuerdo con la definición 3.53, existen n variables aleatorias independientes U_1, \dots, U_n , todas con distribución normal estándar, una matriz de $n \times n$ invertible A y un vector n -dimensional μ tales que $X = AU + \mu$, en donde:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

Sea ahora $\{k_1, \dots, k_n\}$ una permutación de los primeros n números naturales y definamos:

$$\bar{X}' = \begin{pmatrix} X_{k_1} \\ X_{k_2} \\ \vdots \\ X_{k_n} \end{pmatrix}, \quad \bar{U}' = \begin{pmatrix} U_{k_1} \\ U_{k_2} \\ \vdots \\ U_{k_n} \end{pmatrix}$$

A' la matriz formada permutando, de manera similar, los renglones de la matriz A , y μ' el vector obtenido permutando las coordenadas del vector μ , entonces la matriz A' es invertible y $\bar{X}' = A'\bar{U}' + \mu'$, así que la familia de variables aleatorias X_{k_1}, \dots, X_{k_n} también tiene distribución normal multivariada.

Por lo tanto, para demostrar que cualquier subfamilia X_{j_1}, \dots, X_{j_k} tiene distribución normal multivariada, basta con probar que, para cualquier $k \in \{1, \dots, n\}$, la familia X_1, \dots, X_k tiene distribución normal multivariada, lo cual se hace a continuación:

PROPOSICIÓN 3.64. *Sea X un vector aleatorio n -dimensional con distribución normal multivariada y $k \in \{1, \dots, n\}$, entonces el vector k -dimensional formado por las primeras k coordenadas de X tiene distribución normal multivariada.*

Demostración

Si μ es el vector de esperanzas de X , entonces $X - \mu$ también tiene distribución normal multivariada. Si el resultado es válido para $X - \mu$ entonces el vector k -dimensional formado por las primeras k coordenadas de $X - \mu$ tiene distribución normal multivariada; por lo tanto, el vector k -dimensional formado por las primeras k coordenadas de X también tiene distribución normal multivariada. Así que basta con considerar el caso en que el vector de esperanzas de X es cero.

Sabemos que existe un vector aleatorio n -dimensional U formado por variables aleatorias independientes con distribución normal estándar y una matriz de $n \times n$ invertible A tales que $X = AU$.

Consideremos la transformación $V = BX$, en donde B es una matriz de la forma:

$$B = \begin{pmatrix} I_k & 0_{k(n-k)} \\ F & I_{n-k} \end{pmatrix}$$

en donde F es una matriz de $(n-k) \times k$.

El determinante de B es ± 1 , así que B es invertible.

Se tiene $V = BX = (BA)U$, así que V tiene distribución normal multivariada con vector de esperanzas 0 y matriz de covarianzas $C_V = (BA)(BA)^t = BAA^tB^t = BC_XB^t$, en donde C_X es la matriz de covarianzas de X .

Expresemos C_X en la forma siguiente:

$$C_X = \begin{pmatrix} C_k & D \\ D^t & E \end{pmatrix}$$

en donde C_k es una matriz de $k \times k$, D una matriz de $k \times (n-k)$ y E una matriz de $(n-k) \times (n-k)$.

Obviamente, C_k es simétrica. Además, si $z \in \mathbb{R}^k$ es un vector distinto de cero y $w \in \mathbb{R}^n$ es un vector cuyas primeras coordenadas coinciden con las de z y el resto son nulas, entonces $z^t C_k z = w^t C_X w > 0$ ya que C_X es definida positiva, así que C_k también es definida positiva y, por lo tanto, invertible. Además:

$$\begin{aligned} C_V = BC_XB^t &= \begin{pmatrix} I_k & 0_{k(n-k)} \\ F & I_{n-k} \end{pmatrix} \begin{pmatrix} C_k & D \\ D^t & E \end{pmatrix} \begin{pmatrix} I_k & F^t \\ 0_{(n-k)k} & I_{n-k} \end{pmatrix} \\ &= \begin{pmatrix} C_k & C_k F^t + D \\ FC_k + D^t & FC_k F^t + D^t F^t + FD + E \end{pmatrix} \end{aligned}$$

Tomemos F tal que $FC_k + D^t = 0_{k(n-k)}$, es decir, $F = -D^t C_k^{-1}$, entonces:

$$C_V = \begin{pmatrix} C_k & 0_{(n-k)k} \\ 0_{k(n-k)} & E - D^t C_k^{-1} D \end{pmatrix}$$

Como C_V es simétrica y definida positiva, la matriz $G = E - D^t C_k^{-1} D$ también es simétrica y definida positiva.

Por otra parte, Si $X_{(k)}$ es el vector formado por las primeras k coordenadas de X y $X_{(n-k)}$ el vector formado por sus últimas $n - k$ coordenadas, se tiene:

$$V = BX = \begin{pmatrix} I_k & 0_{k(n-k)} \\ F & I_{n-k} \end{pmatrix} \begin{pmatrix} X_{(k)} \\ X_{(n-k)} \end{pmatrix} = \begin{pmatrix} X_{(k)} \\ FX_{(k)} + X_{(n-k)} \end{pmatrix}$$

Es decir, las primeras coordenadas de V y de X coinciden.

Sea v un vector n dimensional, $v_{(k)}$ es el vector formado por las primeras k coordenadas de v y $v_{(n-k)}$ el vector formado por sus últimas $n - k$ coordenadas. Entonces:

$$\begin{aligned} f_V(v) &= \frac{\sqrt{|C_V^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} v^t C_V^{-1} v \right\} \\ &= \frac{\sqrt{|C_k^{-1}| |G^{-1}|}}{(\sqrt{2\pi})^n} \exp \left\{ -\frac{1}{2} \begin{pmatrix} v_{(k)} & v_{(n-k)} \end{pmatrix} \begin{pmatrix} C_k^{-1} & 0_{k(n-k)} \\ 0_{(n-k)k} & G^{-1} \end{pmatrix} \begin{pmatrix} v_{(k)} \\ v_{(n-k)} \end{pmatrix} \right\} \\ &= \frac{\sqrt{|C_k^{-1}|}}{(\sqrt{2\pi})^k} \frac{\sqrt{|G^{-1}|}}{(\sqrt{2\pi})^{n-k}} \exp \left\{ -\frac{1}{2} v_{(k)}^t C_k^{-1} v_{(k)} - \frac{1}{2} v_{(n-k)}^t G^{-1} v_{(n-k)} \right\} \\ &= \frac{\sqrt{|C_k^{-1}|}}{(\sqrt{2\pi})^k} \exp \left\{ -\frac{1}{2} v_{(k)}^t C_k^{-1} v_{(k)} \right\} \frac{\sqrt{|G^{-1}|}}{(\sqrt{2\pi})^{n-k}} \exp \left\{ -\frac{1}{2} v_{(n-k)}^t G^{-1} v_{(n-k)} \right\} \end{aligned}$$

Por lo tanto, una función de densidad conjunta del vector aleatorio $X_{(k)}$ está dada por:

$$f_{X_{(k)}}(v_{(k)}) = \frac{\sqrt{|C_k^{-1}|}}{(\sqrt{2\pi})^k} \exp \left\{ -\frac{1}{2} v_{(k)}^t C_k^{-1} v_{(k)} \right\}$$

Así que, por la proposición 3.56, la familia X_1, X_2, \dots, X_k tiene distribución normal multivariada. ■

3.4. Distribuciones muestrales

De acuerdo con el corolario 3.61, Si U es un vector formado por variables aleatorias independientes, todas con distribución normal estándar, y P es una matriz ortogonal, entonces $V = PU$ también es un vector formado por variables aleatorias independientes, todas con distribución normal estándar. Este resultado se puede extender de la siguiente manera:

PROPOSICIÓN 3.65. *Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución normal de varianza común σ^2 , A una matriz ortogonal de $n \times n$ y μ*

un vector n -dimensional. Definamos las variables aleatorias Y_1, \dots, Y_n mediante la relación $Y = AX + \mu$, en donde:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Entonces las variables aleatorias Y_1, \dots, Y_n son independientes y todas ellas tienen distribución normal con varianza común σ^2 .

Demostración

Para $i \in \{1, \dots, n\}$, sea $\nu_i = E[X_i]$ y sea $\nu = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{pmatrix}$,

entonces las variables aleatorias U_1, \dots, U_n definidas por $U_i = \frac{X_i - \nu_i}{\sigma}$ son independientes y todas tienen distribución normal estándar. Además:

$$Y = AX + \mu = \sigma AU + A\nu + \mu$$

en donde $U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$

Así que, de acuerdo con la proposición 3.61, las variables aleatorias $\frac{1}{\sigma}Y_1, \dots, \frac{1}{\sigma}Y_n$ son independientes y todas tienen distribución normal de varianza 1, de lo cual se sigue el resultado. ■

COROLARIO 3.66. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución normal de varianza común σ^2 y A una matriz ortogonal de $n \times n$. Definamos las variables aleatorias Y_1, \dots, Y_n mediante la relación $Y = AX$, en donde:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

Entonces las variables aleatorias Y_1, \dots, Y_n son independientes, todas ellas tienen distribución normal con varianza común σ^2 y $\sum_{j=1}^n Y_j^2 = \sum_{j=1}^n X_j^2$.

Demostración

Por el corolario 3.65, únicamente resta probar que $\sum_{j=1}^n Y_j^2 = \sum_{j=1}^n X_j^2$.

$$\sum_{j=1}^n Y_j^2 = Y \cdot Y = (AX) \cdot (AX) = (AX)^t(AX) = X^t A^t A X = X^t A^t A X = X \cdot X = \sum_{j=1}^n X_j^2.$$

■

Los resultados anteriores permiten ahora demostrar el siguiente resultado, el cual es de importancia básica en la Estadística.

PROPOSICIÓN 3.67 (Independencia de la media y la varianza muestrales).
Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución $N(\mu, \sigma^2)$. Entonces la media muestral $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ y la varianza muestral $s_X^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$ son independientes.

Demostración

Definamos una nueva familia de variables aleatorias, Y_1, \dots, Y_n , de la siguiente manera (transformación de Helmert):

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k,$$

$$Y_i = \frac{1}{\sqrt{i(i-1)}} \left[\sum_{k=1}^{i-1} X_k - (i-1)X_i \right] \quad \text{para } i \in \{2, \dots, n\}.$$

Sea $A = (a_{ij})$ la matriz de $n \times n$ que transforma el vector $\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$ en el vector

$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$. Se tiene entonces, $a_{ij} = \frac{1}{\sqrt{n}}$, para $j \in \{1, \dots, n\}$ y, para $i \in \{2, \dots, n\}$:

$$a_{ij} = \begin{cases} \frac{1}{\sqrt{i(i-1)}} & \text{si } j \in \{1, \dots, i-1\} \\ -\frac{(i-1)}{\sqrt{i(i-1)}} & \text{si } j = i \\ 0 & \text{si } j \in \{i+1, \dots, n\} \end{cases}$$

Así que, $\sum_{j=1}^n a_{1j}^2 = \sum_{j=1}^n \frac{1}{n} = 1$ y, para $i \in \{2, \dots, n\}$:

$$\sum_{j=1}^n a_{ij}^2 = \sum_{j=1}^n \frac{1}{i(i-1)} + \frac{i-1}{i} = \frac{1}{i} + 1 - \frac{1}{i} = 1$$

También:

$$\sum_{j=1}^n a_{1j}a_{ij} = \frac{1}{\sqrt{n}} \sum_{j=1}^n a_{ij} = \frac{1}{\sqrt{n}} \left[\sum_{j=1}^{i-1} \frac{1}{\sqrt{i(i-1)}} - \frac{(i-1)}{\sqrt{i(i-1)}} \right] = \frac{1}{\sqrt{n}} \left[\frac{i-1}{\sqrt{i(i-1)}} - \frac{(i-1)}{\sqrt{i(i-1)}} \right] = 0$$

y, para $i \in \{2, \dots, n\}$ y $k \in \{3, \dots, n\}$, con $i < k$:

$$\begin{aligned} \sum_{j=1}^n a_{ij}a_{kj} &= \sum_{j=1}^{i-1} \frac{1}{\sqrt{i(i-1)}} \frac{1}{\sqrt{k(k-1)}} - \frac{(i-1)}{\sqrt{i(i-1)}} \frac{1}{\sqrt{k(k-1)}} \\ &= \frac{(i-1)}{\sqrt{i(i-1)}} \frac{1}{\sqrt{k(k-1)}} - \frac{(i-1)}{\sqrt{i(i-1)}} \frac{1}{\sqrt{k(k-1)}} = 0 \end{aligned}$$

De manera que la matriz A es ortogonal. Por lo tanto, de acuerdo con el corolario 3.66, las variables aleatorias Y_1, \dots, Y_n son independientes, cada una de ellas tiene distribución normal de varianza σ^2 y $\sum_{j=1}^n Y_j^2 = \sum_{j=1}^n X_j^2$.

Además:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{k=1}^n X_k = \frac{1}{\sqrt{n}} Y_1 \\ s_X^2 &= \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{k=1}^n X_k^2 - 2\bar{X} \sum_{k=1}^n X_k + n\bar{X}^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{k=1}^n X_k^2 - 2n\bar{X}^2 + n\bar{X}^2 \right] = \frac{1}{n-1} \left[\sum_{k=1}^n X_k^2 - n\bar{X}^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{k=1}^n Y_k^2 - Y_1^2 \right] = \frac{1}{n-1} \sum_{k=2}^n Y_k^2 \end{aligned}$$

De manera que \bar{X} y s_X^2 son variables aleatorias independientes. ■

COROLARIO 3.68. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución $N(\mu, \sigma^2)$. Entonces a) $U = \frac{(n-1)s_X^2}{\sigma^2}$ tiene distribución χ^2 con $n-1$ grados de libertad y b) $V = \frac{\sqrt{n}(\bar{X}-\mu)}{s_X}$ tiene distribución t con $n-1$ grados de libertad.

Demostración

$$\text{a. } U = \frac{(n-1)s_X^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=2}^n Y_k^2 = \sum_{k=2}^n \frac{Y_k^2}{\sigma^2}$$

Pero, para $k \in \{2, \dots, n\}$:

$$\mu_{Y_{ik}} = \frac{1}{\sqrt{k(k-1)}} \left[\sum_{j=1}^{k-1} \mu_{X_j} - (k-1)\mu_{X_k} \right] = \frac{1}{\sqrt{k(k-1)}} \left[\sum_{j=1}^{k-1} \mu - (k-1)\mu \right] = 0$$

Así que $\frac{Y_k}{\sigma}$ tiene distribución normal estándar. Por lo tanto, U tiene distribución \mathcal{X}^2 con $n - 1$ grados de libertad.

$$\text{b. } V = \frac{\sqrt{n}(\bar{X} - \mu)}{s_X} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{U}{n-1}}}$$

Además, $\mu_{Y_1} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \mu_{X_k} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \mu = \sqrt{n}\mu$, así que $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{Y_1 - \sqrt{n}\mu}{\sigma}$ es una variable aleatoria independiente de U y tiene distribución normal estándar. Por lo tanto, V tiene distribución t con $n - 1$ grados de libertad. ■

EJERCICIOS

EJERCICIO 3.1. Sean U y V dos variables aleatorias independientes con distribución normal estándar. Definamos $X = 3U - 4V + 1$ y $Y = 2U + V - 2$. Encuentre la esperanza y la varianza de X y Y , así como el coeficiente de correlación y una función de densidad conjunta de X y Y .

EJERCICIO 3.2. Sea (X, Y) un vector aleatorio con función de densidad $f : \mathbb{R}^2 \mapsto \mathbb{R}$ dada por:

$$f(x, y) = C \exp \left\{ -\frac{1}{2} (2x^2 + 3xy + 5y^2 + x - 4y) \right\}$$

en donde C es una constante.

a) Verifique que el vector (X, Y) tiene distribución normal bivariada.

b) Exprese f en la forma:

$$f(x, y) = K \exp \left\{ -\frac{1}{2} [a(x - \mu)^2 + b(x - \mu)(y - \nu) + c(y - \nu)^2] \right\}$$

en donde K, a, b, c, μ y ν son constantes.

c) Encuentre la esperanza y la varianza de X y Y , así como el coeficiente de correlación entre X y Y .

EJERCICIO 3.3. Sea $X = (X_1, X_2)$ un vector aleatorio con distribución normal bivariada con vector de esperanzas $(\frac{1}{2}, -\frac{1}{2})$, vector de varianzas $(3, 6)$ y coeficiente de correlación $\frac{1}{3}$. Encuentre dos variables aleatorias independientes, U y V , con distribución normal estándar tales que $X = aU + bV + \mu$ y $Y = cU + dV + \nu$, en donde a, b, c, d, μ y ν son constantes

EJERCICIO 3.4. Sea X una variable aleatoria con distribución normal estándar y Z una variable aleatoria, independiente de X , con distribución Bernoulli de parámetro $p = \frac{1}{2}$. Definamos la variable aleatoria Y de la siguiente manera:

$$Y = \begin{cases} X & \text{si } Z = 1 \\ -X & \text{si } Z = 0 \end{cases}$$

Demuestre que a) Y tiene distribución normal estándar, b) X y Y no son independientes y c) $\text{Cov}(X, Y) = 0$. d) ¿Es normal bivariada la distribución conjunta de X y Y ?, ¿por qué?

EJERCICIO 3.5. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 1)$ y coeficiente de correlación ρ . Encuentre $P[X \geq 0, Y \geq 0]$.

EJERCICIO 3.6. Sea $X = (X_1, X_2)$ un vector aleatorio con distribución normal bivariada con vector de esperanzas (μ_1, μ_2) , vector de varianzas (σ^2, σ^2) y coeficiente de correlación $\frac{1}{2}$. Encuentre una función de densidad de $Y = X_1 + X_2$.

EJERCICIO 3.7. Sea $X = (X_1, X_2)$ un vector aleatorio con distribución normal bivariada con vector de esperanzas $(5, 2)$, vector de varianzas $(4, 1)$ y coeficiente de correlación $\frac{1}{3}$. Encuentre una función de densidad de $Y = 2X_1 - 3X_2$.

EJERCICIO 3.8. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 1)$ y coeficiente de correlación $\frac{1}{2}$. Encuentre una función de densidad conjunta de $U = X + 2Y$ y $V = 2X - Y$.

EJERCICIO 3.9. Determine cuáles de las siguientes matrices son invertibles y, en su caso, encuentre su inversa:

$$a) A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 0 & 2 \\ 1 & 3 & 1 \end{pmatrix}$$

$$b) B = \begin{pmatrix} 0 & -1 & \frac{2}{5} & \frac{3}{5} \\ -2 & -\frac{1}{5} & 0 & 1 \\ 1 & 0 & -\frac{2}{5} & \frac{1}{5} \\ -1 & \frac{3}{5} & 0 & 2 \end{pmatrix}$$

$$c) C = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 1 & 0 & -\frac{1}{4} \\ \frac{1}{2} & 0 & 1 & -1 & \frac{3}{2} \\ -\frac{1}{2} & 0 & 3 & 1 & -1 \\ 2 & -1 & \frac{5}{2} & -1 & \frac{1}{2} \\ 3 & -1 & \frac{1}{2} & -3 & 3 \end{pmatrix}$$

$$d) D = \begin{pmatrix} 1 & -1 & 2 & 1 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

EJERCICIO 3.10. *Expresándolas como sumas de cuadrados, determine cuáles de las siguientes formas cuadráticas son definidas positivas:*

$$a) F(x_1, x_2) = 3x_1^2 + x_2^2 + 4x_1x_2$$

$$b) F(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 - x_1x_3$$

$$c) F(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4$$

$$d) F(x_1, x_2, x_3, x_4, x_5) = 2x_1^2 + x_2^2 + x_3^2 + 3x_4^2 + 3x_1x_2 - x_1x_3 + x_2x_4 + x_2x_5 - 2x_4x_5$$

EJERCICIO 3.11. *Determine cuáles de las siguientes matrices Q son definidas positivas:*

$$a) Q = \begin{pmatrix} 3 & -2 \\ -2 & 5 \end{pmatrix}$$

$$b) Q = \begin{pmatrix} 4 & -2 & 5 \\ -2 & 3 & 6 \\ 5 & 6 & 7 \end{pmatrix}$$

$$c) Q = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 2 & -2 & 2 \\ 1 & -2 & 3 & -3 \\ -1 & 2 & -3 & 4 \end{pmatrix}$$

$$\text{EJERCICIO 3.12. Sea } Q = \begin{pmatrix} \frac{5}{2} & -\frac{1}{2} & 2 \\ -\frac{1}{2} & \frac{5}{2} & 2 \\ 2 & 2 & 0 \end{pmatrix}$$

Encuentre una matriz ortogonal P tal que $P^tQP = D_{\alpha_1, \alpha_2, \alpha_3}$, en donde $\alpha_1, \alpha_2, \alpha_3$ son números reales.

EJERCICIO 3.13. *Utilizando el método de los vectores propios, determine cuáles de las siguientes matrices Q son definidas positivas y, en su caso, encuentre una matriz B tal que $Q = B^2$.*

$$a) Q = \begin{pmatrix} 3 & -2 \\ -2 & 5 \end{pmatrix}$$

$$b) Q = \begin{pmatrix} 4 & -2 & 5 \\ -2 & 3 & 6 \\ 5 & 6 & 7 \end{pmatrix}$$

$$c) Q = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 2 & -2 & 2 \\ 1 & -2 & 3 & -3 \\ -1 & 2 & -3 & 4 \end{pmatrix}$$

EJERCICIO 3.14. Sea Q_n la matriz de $n \times n$ definida por:

$$Q_n = \begin{pmatrix} 1 & c & \cdots & c & c \\ c & 1 & \cdots & c & c \\ \vdots & \vdots & \ddots & c & \vdots \\ c & c & \cdots & 1 & c \\ c & c & \cdots & c & 1 \end{pmatrix}$$

en donde c es una constante.

a) Demuestre que Q_n es definida positiva si y sólo si $-\frac{1}{n-1} < c < 1$.

Sugerencia: Defina $P_n(\lambda) = |Q - \lambda I_n|$ y demuestre que:

$$P_n(\lambda) = c(1 - c - \lambda)^{n-1} + (1 - c - \lambda)P_{n-1}(\lambda)$$

b) Demuestre que:

$$\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

son vectores propios de Q .

EJERCICIO 3.15. Sea C la matriz dada por:

$$C = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1 \end{pmatrix}$$

a) Encuentre una matriz ortogonal P tal que $P^t C P = D_{\alpha_1, \alpha_2, \alpha_3, \alpha_4}$, en donde $\alpha_1, \alpha_2, \alpha_3$ y α_4 son números reales positivos.

b) Encuentre una matriz B tal que $C = B^2$.

c) Encuentre una matriz invertible A tal que la matriz de covarianzas de $X = AU$ sea C , en donde U es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar.

EJERCICIO 3.16. Sea X un vector aleatorio con distribución normal multivariada, con vector de esperanzas 0 y matriz de covarianzas C , en donde:

$$a) C = \frac{1}{30} \begin{pmatrix} 79 & 42 & -51 \\ 42 & 36 & -18 \\ -51 & -18 & 54 \end{pmatrix}$$

$$b) C = \frac{1}{18} \begin{pmatrix} 1547 & -492 & -52 & -111 \\ -492 & 162 & 12 & 36 \\ -52 & 12 & 32 & 6 \\ -111 & 36 & 6 & 18 \end{pmatrix}$$

$$c) C = \frac{1}{45} \begin{pmatrix} 121 & -122 & 158 & 104 & 48 \\ -122 & 175 & -130 & -130 & -60 \\ 158 & -130 & 265 & 130 & 60 \\ 104 & -130 & 130 & 130 & 60 \\ 48 & -60 & 60 & 60 & 90 \end{pmatrix}$$

Encuentre una matriz invertible A tal que $X = AU$, en donde U es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar.

EJERCICIO 3.17. Sea C la matriz dada por:

$$C = \begin{pmatrix} 1 & -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1 \end{pmatrix}$$

Utilizando el método de Lagrange, demuestre que C es definida positiva y encuentre una matriz invertible A tal que la matriz de covarianzas de $X = AU$ sea C , en donde U es un vector aleatorio formado por variables aleatorias independientes, todas con distribución normal estándar.

EJERCICIO 3.18. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por $f_{X,Y}(x,y) = \frac{1}{\pi\sqrt{3}} \exp\left\{-\frac{2}{3}(x^2 + y^2 - xy)\right\}$ para cualquier vector $(x,y) \in \mathbb{R}^2$. a) Encuentre la matriz de covarianzas de $U = X + Y$ y $V = X - Y$. b) ¿Son U y V independientes? Justifique su respuesta. c) Encuentre la matriz de covarianzas de X y Y . d) ¿Son X y Y independientes? Justifique su respuesta.

EJERCICIO 3.19. Sean X y Y dos variables aleatorias con distribución normal estándar. a) Encuentre una función de densidad conjunta de $U = 2X + Y$ y $V = X - Y$ y exprésela en la forma $f_{U,V}(\bar{u}) = k \exp \left\{ -\frac{1}{2} Q(\bar{u} - \bar{\mu}) \cdot (\bar{u} - \bar{\mu}) \right\}$. b) Encuentre directamente la matriz de covarianzas C y muestre que Q es la inversa de C .

EJERCICIO 3.20. Sea X_1, \dots, X_8 una muestra aleatoria de una distribución normal con parámetros μ y $\sigma^2 = 15$ y definamos $\bar{X} = \frac{1}{8} \sum_{k=1}^8 X_k$, $s^2 = \frac{1}{7} \sum_{k=1}^8 (X_k - \bar{X})^2$. Encuentre $P[3 < s^2 < 20]$.

EJERCICIO 3.21. Sean X_1, X_2, X_3 y X_4 cuatro variables aleatorias independientes, todas con distribución normal estándar y definamos las variables aleatorias U y V de la siguiente manera:

$$U = X_1 + X_2 + X_3 + X_4$$

$$V = 4(X_1^2 + X_2^2 + X_3^2 + X_4^2) - (X_1 + X_2 + X_3 + X_4)^2$$

Demuestre que U y V son independientes y encuentre sus funciones de densidad.

Sugerencia: Considere la transformación:

$$Y_1 = \frac{1}{2}(X_1 + X_2 + X_3 + X_4),$$

$$Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$$

$$Y_3 = \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3)$$

$$Y_4 = \frac{1}{\sqrt{12}}(X_1 + X_2 + X_3 - 3X_4)$$

EJERCICIO 3.22. Sean X_1, X_2 y X_3 tres variables aleatorias independientes, las 3 con distribución normal estándar, y definamos:

$$Y_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{\sqrt{2}}X_3,$$

$$Y_2 = -\frac{1}{2}X_1 - \frac{1}{2}X_2 + \frac{1}{\sqrt{2}}X_3,$$

$$Y_3 = -\frac{1}{\sqrt{2}}X_1 + \frac{1}{\sqrt{2}}X_2.$$

Encuentre una función de densidad de $Z = \frac{Y_1}{\sqrt{Y_2^2 + Y_3^2}}$.

EJERCICIO 3.23. Sean X_1, X_2, X_3, X_4 cuatro variables aleatorias independientes, todas con distribución normal estándar y definamos:

$$Y_1 = \frac{1}{2}(X_1 + X_2 + X_3 + X_4)$$

$$Y_2 = \frac{1}{2}(X_1 - X_2 + X_3 - X_4)$$

$$Y_3 = \frac{1}{\sqrt{2}}(X_1 - X_3)$$

$$Y_4 = \frac{1}{\sqrt{2}}(X_2 - X_4)$$

Encuentre una función de densidad conjunta de Y_1, Y_2, Y_3, Y_4 y una función de densidad de $Z = \frac{Y_1^2 + Y_2^2}{Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2}$. Además, identifique esta última.

EJERCICIO 3.24. Sean X y Y dos variables aleatorias de esperanza y varianza finitas y coeficiente de correlación ρ . a) Demuestre que la matriz de covarianzas, C , de la pareja X, Y es definida positiva si y sólo si $\rho^2 \neq 1$. b) Asumiendo $\rho^2 \neq 1$, demuestre que:

$$\begin{aligned} & \frac{\sqrt{|C^{-1}|}}{2\pi} e^{-\frac{1}{2}C^{-1}(x-\mu_x, y-\mu_y) \cdot (x-\mu_x, y-\mu_y)} \\ &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_x^2} + \frac{(y-\mu_Y)^2}{\sigma_y^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\} \end{aligned}$$

en donde μ_X, σ_X y μ_Y, σ_Y son la esperanza y la varianza de X y Y , respectivamente.

EJERCICIO 3.25. Demuestre que si $X = (X_1, \dots, X_n)$ es un vector aleatorio con distribución normal multivariada y c_1, \dots, c_n son constantes, no todas cero, entonces $c_1X_1 + \dots + c_nX_n$ tiene distribución normal.

EJERCICIO 3.26. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución normal estándar. Demuestre que las variables aleatorias $U = \sum_{k=1}^n X_k$ y $V = \sum_{k=1}^n \alpha_k X_k$ son independientes si y sólo si $\sum_{k=1}^n \alpha_k = 0$.

EJERCICIO 3.27. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución normal, X el vector con coordenadas X_1, \dots, X_n , ν un vector n -dimensional y A una matriz de $n \times n$ invertible. Demuestre que la distribución del vector aleatorio $Y = AX + \nu$ es normal multivariada.

EJERCICIO 3.28. Sea (X, Y, Z) un vector aleatorio con distribución normal multivariada con vector de esperanzas $(0, 0, 0)$ y matriz de covarianzas:

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}$$

Encuentre e identifique la distribución de $U = X + Y - Z$.

CAPÍTULO 4

ESPERANZAS CONDICIONALES

No se le puede pedir al rigor más que consolidar las conquistas de la intuición.

Jacques Salomon Hadamard

4.1. Generalización de la definición de probabilidad condicional

Sea A un evento de probabilidad positiva y X una variable aleatoria de esperanza finita. Queremos definir la esperanza condicional de X dada la ocurrencia del evento A , $E[X | A]$.

Si X es discreta y x_1, x_2, \dots son sus posibles valores, entonces:

$$E[X] = \sum_k x_k P[X = x_k]$$

De manera que resulta natural definir:

$$E[X | A] = \sum_k x_k P[X = x_k | A]$$

expresión que se puede escribir en la siguiente forma:

$$\begin{aligned} E[X | A] &= \sum_k x_k P[X = x_k | A] = \sum_k x_k \frac{P(A \cap [X = x_k])}{P(A)} \\ &= \frac{1}{P(A)} \sum_k x_k P(A \cap [X = x_k]) = \frac{1}{P(A)} \sum_k x_k P[X I_A = x_k] = \frac{1}{P(A)} E[X I_A]. \end{aligned}$$

Esta última expresión no depende de la forma que tiene X , así que se puede utilizar para dar la siguiente definición general:

DEFINICIÓN 4.1 (Esperanza condicional dada la ocurrencia de un evento).
Sea A un evento de probabilidad positiva y X una variable aleatoria de esperanza

finita. Se define la esperanza condicional de X dada la ocurrencia del evento A , $E[X | A]$, mediante la fórmula:

$$E[X | A] = \frac{1}{P(A)} E[XI_A]$$

Obsérvese que $E[X | A]$ está bien definida pues $E[|XI_A|] \leq E[|X|] < \infty$.

Obsérvese también que esta definición es una extensión de la definición de probabilidad condicional de un evento B dada la ocurrencia de A , $P(B | A) = \frac{P(B \cap A)}{P(A)}$. En efecto, si $X = I_B$, entonces:

$$E[X | A] = \frac{1}{P(A)} E[I_B I_A] = \frac{1}{P(A)} E[I_{B \cap A}] = \frac{P(B \cap A)}{P(A)} = P(B | A)$$

EJEMPLO 4.2. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre a) $E[X | X > 4]$ y b) $E[X | X > Y]$.

Solución

$$a. P[X > 4] = \sum_{x=5}^{\infty} p(1-p)^x = (1-p)^5$$

$$E[XI_{\{X>4\}}] = \sum_{x=5}^{\infty} x f_X(x) = \sum_{x=5}^{\infty} xp(1-p)^x = \frac{4p+1}{p} (1-p)^5$$

$$\text{Así que, } E[X | X > 4] = \frac{4p+1}{p} = \frac{1}{p} + 4.$$

$$P[X = x | X > 4] = \begin{cases} p(1-p)^{x-5} & \text{si } x \in \{5, 6, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

$$b. P[X > Y] = \sum_{y=0}^{\infty} P[X > Y, Y = y] = \sum_{y=0}^{\infty} P[X > y, Y = y]$$

$$= \sum_{y=0}^{\infty} P[X > y] P[Y = y] = \sum_{y=0}^{\infty} (1-p)^{y+1} p(1-p)^y = \frac{1-p}{2-p}.$$

$$E[XI_{\{X>Y\}}] = \sum_{x=0}^{\infty} \sum_{y=0}^{x-1} x f_{X,Y}(x, y) = \sum_{x=0}^{\infty} x P[X = x] P[Y \leq x-1]$$

$$= \sum_{x=0}^{\infty} xp(1-p)^x [1 - (1-p)^x] = \sum_{x=0}^{\infty} xp(1-p)^x - \sum_{x=0}^{\infty} xp(1-p)^{2x}$$

$$= \frac{1-p}{p} - \frac{(1-p)^2}{p(2-p)^2} = \frac{(1-p)(p^2-3p+3)}{p(2-p)^2}.$$

$$\text{Así que, } E[X | X > Y] = \frac{2-p}{1-p} \frac{(1-p)(p^2-3p+3)}{p(2-p)^2} = \frac{p^2-3p+3}{p(2-p)}.$$

EJEMPLO 4.3. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial, X de parámetro λ_1 y Y de parámetro λ_2 . Encuentre:

$$a) E[X | X > 4]$$

$$b) E[X | X > Y]$$

$$c) E[X | X < Y]$$

Solución

$$a. P(X > 4) = \int_4^\infty \lambda_1 e^{-\lambda_1 x} dx = e^{-4\lambda_1}$$

$$E[XI_{[X>4]}] = \int_4^\infty \lambda_1 x e^{-\lambda_1 x} dx = \frac{1}{\lambda_1} (4\lambda_1 + 1) e^{-4\lambda_1}$$

$$\text{Así que, } E[X | X > 4] = \frac{1}{\lambda_1} (4\lambda_1 + 1) = \frac{1}{\lambda_1} + 4.$$

$$b. P(X > Y) = \int_0^\infty \int_0^x \lambda_1 \lambda_2 e^{-\lambda_1 x} e^{-\lambda_2 y} dy dx = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

$$E[XI_{[X>Y]}] = \int_0^\infty \int_0^x \lambda_1 \lambda_2 x e^{-\lambda_1 x} e^{-\lambda_2 y} dy dx = \frac{\lambda_2(2\lambda_1 + \lambda_2)}{\lambda_1(\lambda_1 + \lambda_2)^2}.$$

$$\text{Así que, } E[X | X > Y] = \frac{\lambda_2(2\lambda_1 + \lambda_2)}{\lambda_1(\lambda_1 + \lambda_2)^2} \frac{\lambda_1 + \lambda_2}{\lambda_2} = \frac{2\lambda_1 + \lambda_2}{\lambda_1(\lambda_1 + \lambda_2)} = \frac{1}{\lambda_1} + \frac{1}{\lambda_1 + \lambda_2}.$$

$$c. P(X < Y) = 1 - P(X > Y) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

$$E[XI_{[X<Y]}] = E[X] - E[XI_{[X>Y]}] = \frac{1}{\lambda_1} - \frac{\lambda_2(2\lambda_1 + \lambda_2)}{\lambda_1(\lambda_1 + \lambda_2)^2} = \frac{\lambda_1}{(\lambda_1 + \lambda_2)^2}.$$

$$\text{Así que, } E[X | X < Y] = \frac{\lambda_1}{(\lambda_1 + \lambda_2)^2} \frac{\lambda_1 + \lambda_2}{\lambda_1} = \frac{1}{\lambda_1 + \lambda_2}.$$

▲

Obsérvese que:

$$E[X | X < Y] = \frac{1}{P[X < Y]} E[XI_{[X < Y]}] = \frac{1}{P[X < Y]} \iint_{\{(x,y) \in \mathbb{R}^2: x < y\}} x f_{X,Y}(x,y) dx dy$$

4.2. Esperanzas condicionales en el caso discreto

Sea X una variable aleatoria discreta de esperanza finita y Y cualquier variable aleatoria discreta. Si y es un número real tal que $P[Y = y] > 0$ y x_1, x_2, \dots son los posibles valores de X , se tiene:

$$E[X | Y = y] = \frac{1}{P[Y=y]} E[XI_{[Y=y]}] = \sum_k x_k \frac{f_{X,Y}(x_k, y)}{f_Y(y)}$$

Obsérvese que $E[X | Y = y]$ se calcula de la misma manera que $E[X]$, reemplazando la función de densidad $f_x(x_k)$ de X , por el cociente $\frac{f_{X,Y}(x_k, y)}{f_Y(y)}$. Esto motiva las siguientes definiciones:

Sean X y Y dos variables aleatorias discretas con función de densidad conjunta $f_{X,Y}$. Para cada $y \in \mathbb{R}$ definamos la función $x \mapsto f_{X|Y}(x | y)$ de la siguiente manera:

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ f_X(x) & \text{en otro caso} \end{cases}$$

Obsérvese que, para fines de la definición de $f_{X|Y}$, no importan los valores de $f_{X|Y}(x | y)$ en los puntos y en los cuales $P[Y = y] = 0$ pues Y no toma esos valores.

Obsérvese que, para cualquier $y \in \mathbb{R}$, vista como función de x , $f_{X|Y}(x | y)$ es una función de densidad discreta y, además, si $f_Y(y) > 0$, para cualquier $x \in \mathbb{R}$, se tiene:

$$P[X = x | Y = y] = f_{X|Y}(x | y)$$

Resulta entonces natural definir a la función $x \mapsto f_{X|Y}(x | y)$ como la **función de densidad condicional de X dado que $Y = y$** . Como toda función de densidad, la función de densidad condicional de una variable aleatoria X define una distribución, la cual será llamada la **distribución condicional de X dada Y** .

DEFINICIÓN 4.4 (Esperanza condicional de una variable aleatoria dada otra variable aleatoria - caso discreto). Sea X una variable aleatoria discreta de esperanza finita y Y cualquier variable aleatoria discreta. Definamos la función $h : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:

$$h(y) = \sum_{\{x \in V_X\}} x f_{X|Y}(x | y)$$

en donde V_X es el conjunto de posibles valores de X .

La variable aleatoria $h(Y)$ es llamada la esperanza condicional de X dada Y y se denota por $E[X | Y]$.

Obsérvese que, para fines de la definición de $E[X | Y]$, no importan los valores de h en los puntos y en los cuales $P[Y = y] = 0$ pues Y no toma esos valores.

EJEMPLO 4.5. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{N(N+1)} & \text{si } x \leq y, x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo.

Encuentre a) $E[X | Y]$ y b) $E[Y | X]$.

Solución

a. Para $y \in \{1, \dots, N\}$, se tiene:

$$f_Y(y) = \sum_{x=1}^N f_{X,Y}(x, y) = \sum_{x=1}^y \frac{2}{N(N+1)} = \frac{2}{N(N+1)}y.$$

$$E[X | Y = y] = \sum_{x=1}^N x \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{1}{y} \sum_{x=1}^y x = \frac{1}{2}(y + 1).$$

Así que, $E[X | Y] = \frac{1}{2}(Y + 1)$.

b. Para $x \in \{1, \dots, N\}$ se tiene:

$$f_X(x) = \sum_{y=1}^N f_{X,Y}(x, y) = \sum_{y=x}^N \frac{2}{N(N+1)} = \frac{2}{N(N+1)}(N + 1 - x).$$

$$E[Y | X = x] = \sum_{y=1}^N y \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{1}{N+1-x} \sum_{y=x}^N y = \frac{1}{2}(x + N).$$

Así que, $E[Y | X] = \frac{1}{2}(X + N)$.

EJEMPLO 4.6. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre a) $E[X | Y - X]$ y b) $E[Y | Y - X]$.

Solución

Para $z \in \mathbb{Z}$, se tiene:

$$\begin{aligned} P[Y - X = z] &= P[Y = X + z] = \sum_{x=0}^{\infty} P[Y = X + z, X = x] \\ &= \sum_{x=0}^{\infty} P[Y = x + z] P[X = x] = \begin{cases} p^2(1-p)^z \sum_{x=0}^{\infty} (1-p)^{2x} & \text{si } z \geq 0 \\ p^2(1-p)^z \sum_{x=-z}^{\infty} (1-p)^{2x} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} \frac{p(1-p)^z}{2-p} & \text{si } z \geq 0 \\ \frac{p(1-p)^{-z}}{2-p} & \text{si } z < 0 \end{cases} \end{aligned}$$

$$\begin{aligned} \text{a. } E[X | Y - X = z] &= \sum_{k=0}^{\infty} k \frac{P[X=k, Y-X=z]}{P[Y-X=z]} = \sum_{k=0}^{\infty} k \frac{P[X=k, Y=k+z]}{P[Y-X=z]} \\ &= \begin{cases} p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p)(1-p)^{2z} \sum_{k=-z}^{\infty} k(1-p)^{2k} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p)(1-p)^{2z} \sum_{k=0}^{\infty} (k-z)(1-p)^{2k-2z} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} - zp(2-p) \sum_{k=0}^{\infty} (1-p)^{2k} & \text{si } z < 0 \end{cases} \end{aligned}$$

$$= \begin{cases} \frac{(1-p)^2}{p(2-p)} & \text{si } z \geq 0 \\ \frac{(1-p)^2}{p(2-p)} - z & \text{si } z < 0 \end{cases}$$

Por lo tanto, $E[X | Y - X] = \frac{(1-p)^2}{p(2-p)} - (Y - X)I_{(-\infty, 0)}(Y - X)$.

$$\begin{aligned} b. E[Y | Y - X = z] &= \sum_{k=0}^{\infty} k \frac{P[Y=k, Y-X=z]}{P[Y-X=z]} = \sum_{k=0}^{\infty} k \frac{P[Y=k, X=k-z]}{P[Y-X=z]} \\ &= \begin{cases} p(2-p)(1-p)^{-2z} \sum_{k=z}^{\infty} k(1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} p(2-p) \sum_{k=0}^{\infty} (k+z)(1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} + zp(2-p) \sum_{k=0}^{\infty} (1-p)^{2k} & \text{si } z \geq 0 \\ p(2-p) \sum_{k=0}^{\infty} k(1-p)^{2k} & \text{si } z < 0 \end{cases} \\ &= \begin{cases} \frac{(1-p)^2}{p(2-p)} + z & \text{si } z \geq 0 \\ \frac{(1-p)^2}{p(2-p)} & \text{si } z < 0 \end{cases} \end{aligned}$$

Por lo tanto, $E[Y | Y - X] = \frac{(1-p)^2}{p(2-p)} + (Y - X)I_{[0, \infty)}(Y - X)$.

▲

La siguiente proposición caracteriza a la variable aleatoria $E[X | Y]$ y esta caracterización constituye la base para la definición general de la esperanza condicional en la siguiente sección.

PROPOSICIÓN 4.7. *Sea X una variable aleatoria discreta de esperanza finita y Y cualquier variable aleatoria discreta. La variable aleatoria $h(Y) = E[X | Y]$ tiene esperanza finita y $E[f(Y)h(Y)] = E[f(Y)X]$ para cualquier función $f : \mathbb{R} \mapsto \mathbb{R}$ acotada. Además, si h_1 y h_2 son dos funciones con estas mismas dos propiedades, entonces $P[h_1(Y) = h_2(Y)] = 1$.*

Demostración

Demostremos primero que $h(Y)$ tiene esperanza finita. En efecto, si y_1, y_2, \dots son los posibles valores de Y y x_1, x_2, \dots son los posibles valores de X , se tiene:

$$\begin{aligned} \sum_j |h(y_j)| P[Y = y_j] &\leq \sum_j \sum_k |x_k| \frac{P[X=x_k, Y=y_j]}{P[Y=y_j]} P[Y = y_j] \\ &= \sum_j \sum_k |x_k| P[X = x_k, Y = y_j] = \sum_k |x_k| P[X = x_k] = E[|X|] < \infty \end{aligned}$$

Sea ahora $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función acotada, se tiene entonces:

$$\begin{aligned}
E[f(Y)h(Y)] &= \sum_j f(y_j)h(y_j)P[Y = y_j] \\
&= \sum_j f(y_j) \sum_k x_k \frac{P[X=x_k, Y=y_j]}{P[Y=y_j]} P[Y = y_j] \\
&= \sum_j f(y_j) \sum_k x_k P[X = x_k, Y = y_j] \\
&= \sum_{j,k} f(y_j)x_k P[X = x_k, Y = y_j] = E[f(Y)X]
\end{aligned}$$

Supongamos ahora que $g : \mathbb{R} \mapsto \mathbb{R}$ es una función tal que $g(Y)$ tiene esperanza finita y $E[f(Y)g(Y)] = E[f(Y)X]$ para cualquier función $f : \mathbb{R} \mapsto \mathbb{R}$ acotada. Se tiene entonces:

$$\sum_j f(y_j)g(y_j)P[Y = y_j] = \sum_{j,k} f(y_j)x_k P[X = x_k, Y = y_j]$$

para cualquier función f acotada. En particular, si $f = I_{\{y_j\}}$, se tiene:

$$g(y_j)P[Y = y_j] = \sum_k x_k P[X = x_k, Y = y_j]$$

Así que:

$$g(y_j) = \frac{1}{P[Y=y_j]} \sum_k x_k P[X = x_k, Y = y_j] = \sum_k x_k \frac{P[X=x_k, Y=y_j]}{P[Y=y_j]} = h(y_j)$$

Es decir, $g(y) = h(y)$ para cualquier $y \in \mathbb{R}$ tal que $P[Y = y] > 0$.

De manera que, si h_1 y h_2 satisfacen las dos propiedades mencionadas, $h_1(y) = h_2(y)$ para cualquier y tal que $P[Y = y] > 0$, lo cual implica $P[h_1(Y) = h_2(Y)] = 1$. ■

La última proposición nos dice que la variable aleatoria $h(Y)$ queda caracterizada por las dos propiedades mencionadas. Se puede concluir entonces que se puede definir $E[X | Y] = h(Y)$ mediante cualquier función $h : \mathbb{R} \mapsto \mathbb{R}$ tal que $h(Y)$ tenga esperanza finita y que satisfaga $E[f(Y)h(Y)] = E[f(Y)X]$ para cualquier función $f : \mathbb{R} \mapsto \mathbb{R}$ acotada.

EJEMPLO 4.8. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre $E[XY | X + Y]$.

Solución

Se busca una función h tal que $E[f(X + Y)h(X + Y)] = E[f(X + Y)XY]$ para cualquier función f acotada. Es decir:

$$\begin{aligned}
\sum_{z=0}^{\infty} f(z)h(z)(z+1)p^2(1-p)^z &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} f(j+k)jkp^2(1-p)^{j+k} \\
&= \sum_{k=0}^{\infty} \sum_{z=k}^{\infty} f(z)(z-k)kp^2(1-p)^z = \sum_{z=0}^{\infty} \sum_{k=0}^z f(z)(z-k)kp^2(1-p)^z
\end{aligned}$$

Así que:

$$h(z) = \frac{\sum_{k=0}^z (z-k)kp^2(1-p)^z}{(z+1)p^2(1-p)^z} = \frac{\sum_{k=0}^z (z-k)k}{z+1} = \frac{\frac{1}{6}z(z-1)(z+1)}{z+1} = \frac{1}{6}(z^2 - z)$$

Por lo tanto:

$$E[XY | X + Y] = \frac{1}{6}[(X + Y)^2 - (X + Y)]$$

4.3. Definición general de la esperanza condicional

Las propiedades que caracterizan a la esperanza condicional en el caso discreto motivan la siguiente definición general:

DEFINICIÓN 4.9 (Esperanza condicional de una variable aleatoria dada otra variable aleatoria). *Sea X una variable aleatoria de esperanza finita y Y cualquier variable aleatoria. Si existe una función boreliana $h : \mathbb{R} \mapsto \overline{\mathbb{R}}$ tal que $h(Y)$ es una variable aleatoria de esperanza finita y $E[f(Y)h(Y)] = E[f(Y)X]$ para cualquier función boreliana acotada $f : \mathbb{R} \mapsto \mathbb{R}$, entonces se dice que $h(Y)$ es una versión de la esperanza condicional $E[X | Y]$ y se define $E[X | Y] = h(Y)$ y $E[X | Y = y] = h(y)$ para cualquier $y \in \mathbb{R}$.*

PROPOSICIÓN 4.10. *Sea X una variable aleatoria de esperanza finita, Y cualquier variable aleatoria y $h : \mathbb{R} \mapsto \overline{\mathbb{R}}$ una función boreliana tal que $h(Y)$ es una variable aleatoria de esperanza finita. Entonces $h(Y)$ es una versión de la esperanza condicional $E[X | Y]$ si y sólo si $E[I_B(Y)h(Y)] = E[I_B(Y)X]$ para cualquier conjunto boreliano B de \mathbb{R} .*

Demostración

Si $h(Y)$ es una versión de la esperanza condicional $E[X | Y]$ entonces $E[f(Y)h(Y)] = E[f(Y)X]$ para cualquier función boreliana acotada $f : \mathbb{R} \mapsto \mathbb{R}$, así que, en particular, $E[I_B(Y)h(Y)] = E[I_B(Y)X]$ para cualquier conjunto boreliano B de \mathbb{R} .

Supongamos ahora que $E[I_B(Y)h(Y)] = E[I_B(Y)X]$ para cualquier conjunto boreliano B de \mathbb{R} y sea $f : \mathbb{R} \mapsto \mathbb{R}$ una función boreliana acotada no negativa.

Sabemos que existe una sucesión no decreciente f_n de funciones no negativas de la forma $f_n = \sum_{k=1}^{m_n} a_{nk}I_{A_{nk}}$, en donde A_{nk} es un conjunto boreliano de \mathbb{R} y a_{nk} es un número real positivo. Para cada $n \in \mathbb{N}$, se tiene:

$$\begin{aligned} E[f_n(Y)h(Y)] &= E\left[\sum_{k=1}^{m_n} a_{nk}I_{A_{nk}}(Y)h(Y)\right] \\ &= \sum_{k=1}^{m_n} a_{nk}E[I_{A_{nk}}(Y)h(Y)] = \sum_{k=1}^{m_n} a_{nk}E[I_{A_{nk}}(Y)X] \end{aligned}$$

$$= E \left[\sum_{k=1}^{m_n} a_{nk} I_{A_{nk}}(Y) X \right] = E [f_n(Y) X]$$

Por lo tanto:

$$E [f_n(Y) h^+(Y)] - E [f_n(Y) h^-(Y)] = E [f_n(Y) X^+] - E [f_n(Y) X^-]$$

Tomando límites, se obtiene:

$$\begin{aligned} E [f(Y) h(Y)] &= E [f(Y) h^+(Y)] - E [f(Y) h^-(Y)] \\ &= E [f(Y) X^+] - E [f(Y) X^-] = E [f(Y) X] \end{aligned}$$

Si $f : \mathbb{R} \mapsto \mathbb{R}$ es una función boreliana acotada, entonces se tiene:

$$\begin{aligned} E [f(Y) h(Y)] &= E [f^+(Y) h(Y)] - E [f^-(Y) h(Y)] \\ &= E [f^+(Y) X] - E [f^-(Y) X] = E [f(Y) X] \end{aligned}$$

■

Lo que demostramos en la sección anterior es que, en el caso en que X y Y sean variables aleatorias discretas y X tenga esperanza finita, existe una versión de la esperanza condicional $E[X | Y]$, a saber, $h(Y)$, en donde h es la función definida por $h(y) = \sum_k x_k f_{X|Y}(x | y)$, en donde x_1, x_2, \dots son los posibles valores de X y $f_{X|Y}$ es la función de densidad condicional de X dada Y .

La siguiente proposición generaliza este resultado:

PROPOSICIÓN 4.11. *Sean X y Y dos variables aleatorias discretas y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función boreliana tal que $g(X, Y)$ tiene esperanza finita. Definamos la función:*

$$h(y) = \sum_k g(x_k, y) f_{X|Y}(x_k | y)$$

en donde x_1, x_2, \dots son los posibles valores de X . Entonces $h(Y)$ es una versión de la esperanza condicional $E[g(X, Y) | Y]$.

Demostración

Demostremos primero que h está bien definida. En efecto, si y_1, y_2, \dots son los posibles valores de Y , se tiene:

$$\sum_j \sum_k |g(x_k, y_j)| P[X = x_k, Y = y_j] = E[g(X, Y)] < \infty$$

Por lo tanto, $\sum_k |g(x_k, y)| P[X = x_k, Y = y] < \infty$ para cualquier $y \in \mathbb{R}$.

De manera que, si $P[Y = y] > 0$:

$$\begin{aligned} \sum_k |g(x_k, y)| f_{X|Y}(x_k | y) &= \sum_k |g(x_k, y)| \frac{P[X=x_k, Y=y]}{P[Y=y]} \\ &= \frac{1}{P[Y=y]} \sum_k |g(x_k, y)| P[X = x_k, Y = y] < \infty \end{aligned}$$

Demostremos ahora que la variable aleatoria $h(Y)$ tiene esperanza finita. En efecto, se tiene:

$$\begin{aligned} E[h(Y)] &= \sum_j |h(y_j)| P[Y = y_j] \leq \sum_j \sum_k |g(x_k, y_j)| \frac{P[X=x_k, Y=y_j]}{P[Y=y_j]} P[Y = y_j] \\ &= \sum_j \sum_k |g(x_k, y_j)| P[X = x_k, Y = y_j] = E[g(X, Y)] < \infty \end{aligned}$$

Sea ahora $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función acotada. Se tiene entonces:

$$\begin{aligned} E[f(Y)h(Y)] &= \sum_j f(y_j)h(y_j)P[Y = y_j] \\ &= \sum_j f(y_j) \sum_k g(x_k, y_j)P[X = x_k, Y = y_j] \\ &= \sum_{j,k} f(y_j)g(x_k, y_j)P[X = x_k, Y = y_j] = E[f(Y)g(X, Y)] \end{aligned}$$

■

Obsérvese que, si $P[Y = y] > 0$, el valor de $E[g(X, Y) | Y = y] = h(y)$ coincide con el que se obtiene aplicando directamente la definición, es decir:

$$E[g(X, Y) | Y = y] = \frac{1}{P[Y=y]} E[g(X, Y)I_{[Y=y]}]$$

Obsérvese también que como dos versiones de la esperanza condicional $E[g(X, Y) | Y]$ son iguales con probabilidad 1, entonces, si $H(Y)$ es cualquiera de esas versiones, se tiene:

$$H(y) = \sum_k g(x_k, y) \frac{P[X=x_k, Y=y]}{P[Y=y]}$$

para cualquier y tal que $P[Y = y] > 0$.

EJEMPLO 4.12. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre:

a) $E[\min(X, Y) | Y]$

b) $E[\max(X, Y) | Y]$

Solución

a. Para $y \in \{0, 1, \dots\}$, se tiene:

$$\sum_{k=0}^{\infty} \min(k, y) \frac{P[X=k, Y=y]}{P[Y=y]} = \sum_{k=0}^{\infty} \min(k, y) P[X = k]$$

$$\begin{aligned}
&= \sum_{k=0}^y kP[X = k] + \sum_{k=y+1}^{\infty} yP[X = k] \\
&= p \sum_{k=0}^y k(1-p)^k + py \sum_{k=y+1}^{\infty} (1-p)^k \\
&= \frac{1-p}{p} - \frac{py+1}{p}(1-p)^{y+1} + y(1-p)^{y+1} = \frac{1-p}{p} - \frac{1}{p}(1-p)^{y+1}.
\end{aligned}$$

Por lo tanto, $E[\text{mín}(X, Y) | Y] = \frac{1-p}{p} - \frac{1}{p}(1-p)^{Y+1}$.

b. Para $y \in \{0, 1, \dots\}$, se tiene:

$$\begin{aligned}
&\sum_{k=0}^{\infty} \text{máx}(k, y) \frac{P[X=k, Y=y]}{P[Y=y]} = \sum_{k=0}^{\infty} \text{máx}(k, y) P[X = k] \\
&= \sum_{k=0}^{y-1} yP[X = k] + \sum_{k=y}^{\infty} kP[X = k] \\
&= py \sum_{k=0}^{y-1} (1-p)^k + p \sum_{k=y}^{\infty} k(1-p)^k \\
&= y - y(1-p)^y + y(1-p)^y + \frac{1}{p}(1-p)^{y+1} \\
&= y + \frac{1}{p}(1-p)^{y+1}
\end{aligned}$$

Por lo tanto, $E[\text{máx}(X, Y) | Y] = Y + \frac{1}{p}(1-p)^{Y+1}$.

▲

Obsérvese en el último ejemplo que si y es un número real tal que $P[Y = y] > 0$ entonces $E[\text{mín}(X, Y) | Y = y]$ es simplemente la esperanza (no condicional) de la variable aleatoria $\text{mín}(X, y)$. Este resultado se puede generalizar. En efecto, si X y Y son dos variables aleatorias discretas independientes, y un número real tal que $P[Y = y] > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función tal que $g(X, Y)$ tiene esperanza finita, entonces, denotando por x_1, x_2, \dots a los posibles valores de X , se tiene:

$$\begin{aligned}
E[g(X, Y) | Y = y] &= \sum_k g(x_k, y) \frac{P[X=x_k, Y=y]}{P[Y=y]} = \sum_k g(x_k, y) P[X = x_k] \\
&= E[g(X, y)]
\end{aligned}$$

Es decir, la esperanza condicional de $g(X, Y)$, dado que $Y = y$, es simplemente la esperanza (no condicional) de la variable aleatoria $g(X, y)$.

La existencia de una versión de la esperanza condicional de cualquier variable aleatoria de esperanza finita, dada otra variable aleatoria cualquiera, es un resultado que puede probarse. Sin embargo, la demostración general requiere de resultados que rebasan el nivel de este libro. Por tal motivo únicamente se enuncia aquí el resultado general sin prueba.

TEOREMA 4.13. *Sea X una variable aleatoria de esperanza finita y Y cualquier variable aleatoria. Existe entonces una función boreliana $h : \mathbb{R} \mapsto \overline{\mathbb{R}}$ tal que $h(Y)$ es una versión de la esperanza condicional $E[X | Y]$. Además, dos versiones de la esperanza condicional de X con respecto a Y son iguales con probabilidad 1.*

La siguiente proposición muestra que la esperanza condicional tiene propiedades similares a las de la esperanza no condicional. Se muestra también que tiene las propiedades que podrían esperarse con una buena definición, por corresponder a la idea intuitiva del concepto, por ejemplo, una buena definición de la esperanza condicional debería ser tal que si X y Y son independientes entonces el hecho de que Y tome un cierto valor y no debería alterar el valor esperado de X , es decir debería de tenerse $E[X | Y] = E[X]$. Finalmente, se muestran otras propiedades específicas de la esperanza condicional, las cuales no resultan evidentes a partir de la idea intuitiva.

PROPOSICIÓN 4.14. *Sea Y cualquier variable aleatoria. Se tienen entonces las siguientes propiedades:*

- (i) $E[c | Y] = c$ para cualquier constante c .
- (ii) $E[cX | Y] = cE[X | Y]$ para cualquier constante c y cualquier variable aleatoria X de esperanza finita.
- (iii) $E[X_1 + X_2 | Y] = E[X_1 | Y] + E[X_2 | Y]$ para cualquier par de variables aleatorias X_1 y X_2 de esperanza finita.
- (iv) Si X es una variable aleatoria de esperanza finita, entonces:
 $E[E[X | Y]] = E[X]$
- (v) Si X es una variable aleatoria de esperanza finita y $Z = g(Y)$, entonces:
 $E[E(X | Y) | Z] = E(X | Z)$
- (vi) Si X es una variable aleatoria de esperanza finita e independiente de Y , entonces:
 $E[X | Y] = E[X]$
- (vii) Si X es una variable aleatoria no negativa de esperanza finita, entonces, con probabilidad 1, $E[X | Y]$ es no negativa.
- (viii) Si X es una variable aleatoria de esperanza finita y $g : \mathbb{R} \mapsto \mathbb{R}$ es una función boreliana tal que $g(Y)X$ es una variable aleatoria de esperanza finita, entonces $g(Y)E[X | Y]$ tiene esperanza finita y
 $E[g(Y)X | Y] = g(Y)E[X | Y]$
- (ix) Si X es una variable aleatoria de varianza finita, entonces $Z = E[X | Y]$ también tiene varianza finita.

Demostración

Las demostraciones de *i*, *ii* y *iii* se dejan como ejercicio.

iv es un caso particular de una de las propiedades que caracterizan a la esperanza condicional. De manera específica, sabemos que la esperanza condicional $E[X | Y]$ tiene la propiedad de que $E[f(Y)E[X | Y]] = E[f(Y)X]$ para cualquier función f acotada. En particular, considerando la función $f \equiv 1$, se tiene $E[E[X | Y]] = E[X]$.

Para probar *v*, sean $h_1(Y) = E[X | Y]$, $h_2(Z) = E[h_1(Y) | Z]$ y $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana acotada. Se tiene entonces:

$$\begin{aligned} E[f(Z)h_2(Z)] &= E[f(Z)h_1(Y)] = E[(f \circ g)(Y)h_1(Y)] \\ &= E[(f \circ g)(Y)X] = E[f(Z)X] \end{aligned}$$

Por lo tanto, $E[E(X | Y) | Z] = h_2(Z)$ es una versión de $E[X | Z]$.

Para probar *vi*, sea $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana acotada. Se tiene entonces:

$$E[f(Y)X] = E[f(Y)]E[X] = E[f(Y)E[X]]$$

Por lo tanto, $E[X]$ es una versión de $E[X | Y]$.

Para probar *vii*, sea $h(Y) = E[X | Y]$ y, para cada $n \in \mathbb{N}$, definamos:

$$B_n = \left\{ y \in \mathbb{R} : h(y) < -\frac{1}{n} \right\}$$

Entonces:

$$0 \leq E[I_{B_n}(Y)X] = E[I_{B_n}(Y)h(Y)] \leq -\frac{1}{n}P[Y \in B_n] = -\frac{1}{n}P[h(Y) < -\frac{1}{n}]$$

Así que $P[h(Y) < -\frac{1}{n}] = 0$ para cualquier $n \in \mathbb{N}$.

Por lo tanto:

$$P[h(Y) < 0] = \bigcup_{n=1}^{\infty} P[h(Y) < -\frac{1}{n}] = 0$$

Es decir:

$$P[h(Y) \geq 0] = 1$$

Para probar *viii*, supongamos primero que X y g son no negativas y sea Z una versión de la esperanza condicional de X con respecto a Y . Sabemos que existe una sucesión no decreciente g_n de funciones no negativas de la forma $g_n = \sum_{k=1}^{m_n} a_{nk}I_{A_{nk}}$, en donde A_{nk} es un conjunto boreliano de \mathbb{R} y a_{nk} es un número real positivo. Para cada conjunto boreliano B de \mathbb{R} y $n \in \mathbb{N}$, se tiene:

$$E[I_B(Y)g_n(Y)Z] = E[I_B(Y)\sum_{k=1}^{m_n} a_{nk}I_{A_{nk}}(Y)Z]$$

$$\begin{aligned}
&= \sum_{k=1}^{m_n} a_{nk} E [I_B(Y) I_{A_{nk}}(Y) Z] = \sum_{k=1}^{m_n} a_{nk} E [I_B(Y) I_{A_{nk}}(Y) X] \\
&= E [I_B(Y) \sum_{k=1}^{m_n} a_{nk} I_{A_{nk}}(Y) X] = E [I_B(Y) g_n(Y) X]
\end{aligned}$$

Tomando límites, se obtiene:

$$\begin{aligned}
E [I_B(Y) g(Y) Z] &= \lim_{n \rightsquigarrow \infty} E [I_B(Y) g_n(Y) Z] \\
&= \lim_{n \rightsquigarrow \infty} E [I_B(Y) g_n(Y) X] = E [I_B(Y) g(Y) X]
\end{aligned}$$

Así que $g(Y)Z$ tiene esperanza finita y es una versión de $E [g(Y)X | Y]$.

Para el resultado general, se puede escribir

$$\begin{aligned}
g(Y)X &= [g^+(Y) - g^-(Y)] (X^+ - X^-) \\
&= g^+(Y)X^+ + g^-(Y)X^- - g^+(Y)X^- - g^-(Y)X^+
\end{aligned}$$

Además, $(g(Y)X)^+ = g^+(Y)X^+ + g^-(Y)X^-$ y $(g(Y)X)^- = g^+(Y)X^- + g^-(Y)X^+$, así que, como X y $g(Y)X$ tienen esperanza finita, X^+ , X^- , $g^+(Y)X^+$, $g^-(Y)X^-$, $g^+(Y)X^-$ y $g^-(Y)X^+$ también tienen esperanza finita, por lo tanto se puede aplicar, en cada caso, la primera parte de la demostración.

Para probar *ix*, obsérvese que $X^2 - Z^2 \geq 2Z(X - Z)$ y, para cada $n \in \mathbb{N}$, sea $W_n = I_{[-n, n]}(Z)$. Se tiene entonces $W_n X^2 - W_n Z^2 \geq 2W_n Z(X - Z)$. Pero, como W_n y $W_n Z$ son variables aleatorias acotadas que dependen de Y , se tiene:

$$\begin{aligned}
W_n E [X^2 | Y] - W_n Z^2 &= E [W_n X^2 | Y] - E [W_n Z^2 | Y] \\
&\geq 2E [W_n Z(X - Z) | Y] = 2W_n Z E [(X - Z) | Y] = 0
\end{aligned}$$

Así que, para cualquier $n \in \mathbb{N}$:

$$W_n Z^2 \leq W_n E [X^2 | Y]$$

de lo cual se obtiene, tomando límites cuando $n \rightsquigarrow \infty$, $Z^2 \leq E [X^2 | Y]$. Finalmente, tomando esperanzas en ambos miembros de la última desigualdad, se concluye que:

$$E [Z^2] \leq E [X^2] < \infty$$

■

EJEMPLO 4.15. *Sea X una variable aleatoria de esperanza y varianza finitas, Y cualquier variable aleatoria y $Z = E[X | Y]$. Demuestre que a) $E [(X - Z)^2] = E [X^2 - Z^2]$ y b) $\text{Var}(X - Z) = \text{Var}(X) - \text{Var}(Z)$.*

Solución

$$\begin{aligned}
a. \quad E[(X - Z)^2 | Y] &= E[X^2 - 2XZ + Z^2 | Y] \\
&= E[X^2 | Y] - 2ZE[X | Y] + Z^2 \\
&= E[X^2 | Y] - 2Z^2 + Z^2 = E[X^2 - Z^2 | Y]
\end{aligned}$$

Así que, tomando esperanzas:

$$E[(X - Z)^2] = E[X^2 - Z^2]$$

b. Tomando en cuenta que $E[X] = E[Z]$, se tiene:

$$\begin{aligned}
\text{Var}(X - Z) &= E[(X - Z)^2] - (E[X - Z])^2 = E[(X - Z)^2] \\
&= E[X^2] - E[Z^2] = E[X^2] - (E[X])^2 - E[Z^2] + (E[Z])^2 \\
&= \text{Var}(X) - \text{Var}(Z)
\end{aligned}$$

EJEMPLO 4.16. Sean X una variable aleatoria de esperanza y varianza finitas, Y cualquier variable aleatoria y $h(Y) = E[X | Y]$. Demuestre que $E[(X - h(Y))^2] \leq E[(X - g(Y))^2]$ para cualquier función g tal que $g(Y)$ tenga esperanza y varianza finitas.

Solución

Por el inciso a del ejemplo anterior, se tiene:

$$E[(X - h(Y))^2] = E[X^2] - E[h^2(Y)]$$

Por otra parte:

$$\begin{aligned}
E[(X - g(Y))^2 | Y] &= E[X^2 - 2Xg(Y) + g^2(Y) | Y] \\
&= E[X^2 | Y] - 2g(Y)E[X | Y] + g^2(Y) = E[X^2 | Y] - 2g(Y)h(Y) + g^2(Y)
\end{aligned}$$

Así que:

$$E[(X - g(Y))^2] = E[X^2] - 2E[g(Y)h(Y)] + E[g^2(Y)]$$

De manera que:

$$\begin{aligned}
E[(X - g(Y))^2] - E[(X - h(Y))^2] &= E[h^2(Y)] - 2E[g(Y)h(Y)] + E[g^2(Y)] \\
&= E[h^2(Y) - 2g(Y)h(Y) + g^2(Y)] = E[(h(Y) - g(Y))^2] \geq 0
\end{aligned}$$

▲

Obsérvese que el último ejemplo muestra que, cuando se conoce Y , $E[X | Y]$ es un buen estimador de X en el sentido de que, entre todas las funciones g tales que $g(Y)$ tiene esperanza y varianza finitas, $E[X | Y]$ minimiza el valor de $E[(X - g(Y))^2]$. Por tal motivo se puede decir que, conociendo el valor de Y , $E[X | Y]$ es el **mejor estimador de X en el sentido de la media cuadrática**.

4.4. Esperanzas condicionales en el caso absolutamente continuo

Al igual que en el caso discreto, se puede dar la forma explícita de la esperanza condicional en el caso absolutamente continuo. La demostración de que efectivamente se obtiene una versión de ésta requiere de algunos detalles técnicos sobre la integral de Lebesgue. Por tal motivo, primero consideraremos un ejemplo en el cual no se presentan esos detalles técnicos y después demostraremos el resultado general.

EJEMPLO 4.17. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Sea $Z = Y - X$ y definamos:

$$h(z) = \begin{cases} \int_{-\infty}^{\infty} x \frac{f_{X,Z}(x,z)}{f_Z(z)} dx & \text{si } f_Z(z) > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Demuestre que $h(Z)$ es una versión de la esperanza condicional $E[X | Z]$.

Solución

Se tiene:

$$f_{X,Z}(x,z) = f_X(x)f_Y(z+x) = \begin{cases} \lambda^2 e^{-\lambda z} e^{-2\lambda x} & \text{si } x > -z, x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$f_Z(z) = \begin{cases} \int_0^{\infty} \lambda^2 e^{-\lambda z} e^{-2\lambda x} dx & \text{si } z > 0 \\ \int_{-z}^{\infty} \lambda^2 e^{-\lambda z} e^{-2\lambda x} dx & \text{si } z \leq 0 \end{cases} = \begin{cases} \frac{1}{2} \lambda e^{-\lambda z} & \text{si } z > 0 \\ \frac{1}{2} \lambda e^{\lambda z} & \text{si } z \leq 0 \end{cases}$$

$$h(z) = \begin{cases} \int_0^{\infty} 2\lambda x e^{-2\lambda x} dx & \text{si } z > 0 \\ \int_{-z}^{\infty} 2\lambda x e^{-2\lambda(x+z)} dx & \text{si } z \leq 0 \end{cases}$$

$$= \begin{cases} \int_0^{\infty} 2\lambda x e^{-2\lambda x} dx & \text{si } z > 0 \\ \int_0^{\infty} 2\lambda(x-z) e^{-2\lambda x} dx & \text{si } z \leq 0 \end{cases} = \begin{cases} \frac{1}{2\lambda} & \text{si } z > 0 \\ \frac{1}{2\lambda} - z & \text{si } z \leq 0 \end{cases}$$

$$= \frac{1}{2\lambda} - z I_{(-\infty, 0]}(z)$$

$$\text{Así que, } h(Z) = \frac{1}{2\lambda} - Z I_{(-\infty, 0]}(Z).$$

Evidentemente $h(Z)$ es una variable aleatoria. Además:

$$\begin{aligned} \int_{-\infty}^{\infty} |h(z)| f_Z(z) dz &= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} x \frac{f_{X,Z}(x,z)}{f_Z(z)} dx \right| f_Z(z) dz \\ &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| \frac{f_{X,Z}(x,z)}{f_Z(z)} f_Z(z) dx dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| f_{X,Z}(x,z) dx dz \\ &= E[|X|] < \infty \end{aligned}$$

Así que, $h(Z)$ tiene esperanza finita.

Sea ahora $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana acotada. Se tiene entonces:

$$\begin{aligned} E[f(Z)h(Z)] &= \int_{-\infty}^{\infty} f(z)h(z)f_Z(z) dz = \int_{-\infty}^{\infty} f(z) \int_{-\infty}^{\infty} x f_{X,Z}(x,z) dx dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z)x f_{X,Z}(x,z) dx dz E[f(Z)X] \end{aligned}$$

Por lo tanto, $h(Z)$ es una versión de la esperanza condicional $E[X | Z]$. ▲

PROPOSICIÓN 4.18. *Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Para cada $y \in \mathbb{R}$, definamos la función $x \mapsto f_{X|Y}(x | y)$ de la siguiente manera:*

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ f_X(x) & \text{en otro caso} \end{cases}$$

Entonces, para cualquier $y \in \mathbb{R}$, la función $x \mapsto f_{X|Y}(x | y)$ es una función de densidad.

Demostración

La función $x \mapsto f_{X|Y}(x | y)$ es no negativa y se tiene:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X|Y}(x | y) dx &= \begin{cases} \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx & \text{si } f_Y(y) > 0 \\ \int_{-\infty}^{\infty} f_X(x) dx & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{f_Y(y)} f_Y(y) & \text{si } f_Y(y) > 0 \\ \int_{-\infty}^{\infty} f_X(x) dx & \text{en otro caso} \end{cases} = 1 \end{aligned}$$

PROPOSICIÓN 4.19. *Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función boreliana tal que $g(X, Y)$ tiene esperanza finita. Definamos la función $h : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:* ■

$$h(y) = \begin{cases} \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x | y) dx & \text{si } f_Y(y) > 0 \text{ y} \\ & \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dx < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Entonces $h(Y)$ es una versión de la esperanza condicional $E[g(X, Y) | Y]$.

Demostración

Recordemos que, siendo la función $f_{X,Y}$ integrable, por el teorema de Fubini, el conjunto de puntos $y \in \mathbb{R}$ para los cuales la función $x \mapsto f_{X,Y}(x, y)$ no es integrable, tiene medida cero. Además:

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx & \text{si } \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Definamos los siguientes conjuntos.

$$A = \{y \in \mathbb{R} : f_Y(y) > 0\}$$

$$B = \left\{y \in \mathbb{R} : \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dx < \infty\right\}$$

$$C = \left\{y \in \mathbb{R} : \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx < \infty\right\}$$

$$D_y = \{x \in \mathbb{R} : f_{X,Y}(x, y) > 0\}$$

Como mencionamos antes, C^c tiene medida cero. Además:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dx dy < \infty$$

Así que, por el teorema de Fubini, B^c tiene medida cero.

Por otra parte, si $y \in A^c \cap C$, entonces $\int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = 0$, así que D_y tiene medida cero.

También por el teorema de Fubini, h es una función boreliana, así que $h(Y)$ es una variable aleatoria.

Demostremos que $h(Y)$ tiene esperanza finita. En efecto, se tiene:

$$\begin{aligned} \int_{-\infty}^{\infty} |h(y)| f_Y(y) dy &= \int_{A \cap B \cap C} |h(y)| f_Y(y) dy \\ &\leq \int_{A \cap B \cap C} \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dx dy \\ &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dx dy < \infty \end{aligned}$$

Sea $f : \mathbb{R} \mapsto \mathbb{R}$ cualquier función boreliana acotada. Entonces:

$$\begin{aligned}
E[f(Y)h(Y)] &= \int_{-\infty}^{\infty} f(y)h(y)f_Y(y)dy \\
&= \int_{A \cap B \cap C} f(y) \int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy \\
&= \int_{A \cap C} \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy \\
&= \int_{A \cap C} \int_{D_y} f(y)g(x,y)f_{X,Y}(x,y)dxdy \\
&= \int_C \int_{D_y} f(y)g(x,y)f_{X,Y}(x,y)dxdy \\
&= \int_C \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y)g(x,y)f_{X,Y}(x,y)dxdy = E[f(Y)g(X,Y)]
\end{aligned}$$

Así que $h(Y)$ es una versión de la esperanza condicional $E[g(X,Y) | Y]$.

■

Recuérdese que, si $h(Y)$ es una versión de la esperanza condicional $E[X | Y]$, se define $E[X | Y = y] = h(y)$ para cualquier $y \in \mathbb{R}$. De manera que dado el vector aleatorio (X, Y) y la función g de la última proposición, se tiene:

$$E[g(X, Y) | Y = y] = \begin{cases} \int_{-\infty}^{\infty} g(x, y)f_{X|Y}(x | y)dx & \text{si } f_Y(y) > 0 \text{ y} \\ & \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y)dx < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Debe observarse que la definición $h(y) = E[g(X, Y) | Y = y] = 0$ en los puntos $y \notin \{y \in \mathbb{R} : f_Y(y) > 0\} \cap \{y \in \mathbb{R} : \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y)dx < \infty\}$ no tiene influencia sustancial en la definición de $E[g(X, Y) | Y]$ pues la probabilidad de que Y tome valores en ese conjunto es cero. En efecto, si $A = \{y \in \mathbb{R} : f_Y(y) > 0\}$ y $B = \{y \in \mathbb{R} : \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y)dx < \infty\}$, se tiene:

$$P[Y \in A^c] = \int_{A^c} f_Y(y)dy = 0$$

Por otra parte, se tiene:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y)dxdy < \infty$$

Así que, por el teorema de Fubini, B^c tiene medida cero. Por lo tanto:

$$P[Y \in B^c] = \int_{B^c} f_Y(y)dy = 0$$

EJEMPLO 4.20. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda x} & \text{si } 0 < y < x \\ 0 & \text{en otro caso.} \end{cases}$$

Encuentre a) $E[X | Y]$, b) $E[e^{-X} | Y]$, c) $E[e^{-(X-Y)} | Y]$ y d) $E[e^{-XY} | Y]$.

Solución

$$a. f_Y(y) = \begin{cases} \int_y^\infty \lambda^2 e^{-\lambda x} dx & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \lambda e^{-\lambda y} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$E[X | Y = y] = \begin{cases} \int_y^\infty \lambda x e^{-\lambda(x-y)} dx & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} y + \frac{1}{\lambda} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que, $E[X | Y] = Y + \frac{1}{\lambda}$.

$$b. E[e^{-X} | Y = y] = \begin{cases} \int_y^\infty \lambda e^{-x} e^{-\lambda(x-y)} dx & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{\lambda}{\lambda+1} e^{-y} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que, $E[e^{-X} | Y] = \frac{\lambda}{\lambda+1} e^{-Y}$.

c. Las variables aleatorias e^{-X} y $e^{-(X-Y)}$ tienen esperanza finita. Por lo tanto:

$$E[e^{-(X-Y)} | Y] = E[e^{-X} e^Y | Y] = e^Y E[e^{-X} | Y] = e^Y \frac{\lambda}{\lambda+1} e^{-Y} = \frac{\lambda}{1+\lambda}$$

$$d. E[e^{-XY} | Y = y] = \begin{cases} \int_y^\infty \lambda e^{-xy} e^{-\lambda(x-y)} dx & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{\lambda}{\lambda+y} e^{-y^2} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que, $E[e^{-XY} | Y] = \frac{\lambda}{\lambda+Y} e^{-Y^2}$.

EJEMPLO 4.21. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $E[XY | X]$ y b) $E[X | XY]$.

Solución

$$a. E[XY | X] = XE[Y | X] = XE[Y] = \frac{1}{2}X$$

$$b. f_{X,XY}(x, v) = \frac{1}{|x|}f_{X,Y}(x, \frac{v}{x}) = \begin{cases} \frac{1}{x} & \text{si } 0 < v < x < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{XY}(v) = \int_{-\infty}^{\infty} f_{X,XY}(x, v)dx = \begin{cases} \int_v^1 \frac{1}{x} dx & \text{si } 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} -\ln v & \text{si } 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$E[X | XY = v] = \begin{cases} -\int_v^1 \frac{1}{\ln v} dx & \text{si } 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{v-1}{\ln v} & \text{si } 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\text{Así que, } E[X | XY] = \frac{XY-1}{\ln(XY)}.$$

EJEMPLO 4.22. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por $f_{X,Y}(x, y) = 4xy$ si $0 < x < 1$ y $x < y < 1$ ó $-1 < x < 0$ y $x < y < 0$. Encuentre $E[X + Y | Y - X]$.

Solución

$$f_{X+Y, Y-X}(u, v) = \frac{1}{2}f_{X,Y}\left(\frac{u-v}{2}, \frac{u+v}{2}\right)$$

$$= \begin{cases} \frac{1}{2}(u^2 - v^2) & \text{si } v - 2 < u < -v < 0 \text{ ó } 0 < v < u < -v + 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$E[X + Y | Y - X = v]$$

$$= \begin{cases} \frac{1}{2f_{Y-X}(v)} \left(\int_{v-2}^{-v} u(u^2 - v^2) du + \int_v^{-v+2} u(u^2 - v^2) du \right) & \text{si } 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= 0$$

$$\text{Así que, } E[X + Y | Y - X] = 0.$$

EJEMPLO 4.23. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $E[\text{mín}(X, Y) | Y]$ y b) $E[\text{máx}(X, Y) | Y]$.

Solución

$$a. E[\min(X, Y) | Y = y]$$

$$= \begin{cases} \int_0^1 \min(x, y) \frac{f_{X,Y}(x,y)}{f_Y(y)} dx & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \int_0^1 \min(x, y) f_X(x) dx & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \int_0^y x dx + \int_y^1 y dx & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} y - \frac{1}{2}y^2 & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, $E[\min(X, Y) | Y] = Y - \frac{1}{2}Y^2$.

$$b. E[\max(X, Y) | Y] = E[X + Y - \min(X, Y) | Y]$$

$$= E[X] + Y - E[\min(X, Y) | Y]$$

$$= \frac{1}{2} + Y - (Y - \frac{1}{2}Y^2) = \frac{1}{2}(1 + Y^2)$$

▲

Obsérvese en el último ejemplo que si y es un número real tal que $f_Y(y) > 0$ entonces $E[\min(X, Y) | Y = y]$ es simplemente la esperanza (no condicional) de la variable aleatoria $\min(X, y)$. Al igual que en el caso discreto, este resultado se puede generalizar. En efecto, si X y Y son dos variables aleatorias absolutamente continuas independientes, y un número real tal que $f_Y(y) > 0$ y $g: \mathbb{R}^2 \mapsto \mathbb{R}$ una función tal que $g(X, Y)$ y $g(X, y)$ tienen esperanza finita, entonces se tiene:

$$E[g(X, Y) | Y = y] = \int_{-\infty}^{\infty} g(x, y) \frac{f_{X,Y}(x,y)}{f_Y(y)} dx = \int_{-\infty}^{\infty} g(x, y) f_X(x) dx$$

$$= E[g(X, y)].$$

Es decir, la esperanza condicional de $g(X, Y)$, dado que $Y = y$, es simplemente la esperanza (no condicional) de la variable aleatoria $g(X, y)$.

EJEMPLO 4.24. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Encuentre $E[X^n | X^2 + Y^2]$ para cualquier $n \in \mathbb{N}$.

Solución

Busquemos una función h tal que $E[h(X^2 + Y^2)f(X^2 + Y^2)] = E[f(X^2 + Y^2)X^n]$ para cualquier función f acotada.

Como $X^2 + Y^2$ tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$, se tiene:

$$E [h(X^2 + Y^2)f(X^2 + Y^2)] = \frac{1}{2} \int_0^\infty h(z)f(z)e^{-\frac{z}{2}} dz.$$

$$\begin{aligned} E [f(X^2 + Y^2)X^n] &= \frac{1}{2\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty f(x^2 + y^2)x^n e^{-\frac{1}{2}(x^2+y^2)} dx dy \\ &= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(r^2)r^n \cos^n \theta e^{-\frac{1}{2}r^2} r d\theta dr = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} r^{n+1} f(r^2)e^{-\frac{1}{2}r^2} \cos^n \theta d\theta dr \\ &= C \int_0^\infty r^{n+1} f(r^2)e^{-\frac{1}{2}r^2} dr = \frac{C}{2} \int_0^\infty z^{\frac{1}{2}n} f(z)e^{-\frac{1}{2}z} dz \end{aligned}$$

$$\text{en donde } C = \frac{1}{2\pi} \int_0^{2\pi} \cos^n \theta d\theta.$$

Por lo tanto, $h(z) = Cz^{\frac{1}{2}n}$. Es decir:

$$E [X^n | X^2 + Y^2] = C (X^2 + Y^2)^{\frac{n}{2}}$$

Para n impar, se tiene $\int_0^{2\pi} \cos^n \theta d\theta = 0$.

Para n par, se tiene $\int \cos^n \theta d\theta = \frac{1}{n} \cos^{n-1} \theta \sin \theta + \frac{n-1}{n} \int \cos^{n-2} \theta d\theta$. Así que:

$$\int_0^{2\pi} \cos^n \theta d\theta = \frac{n-1}{n} \int_0^{2\pi} \cos^{n-2} \theta d\theta = \dots = \frac{1 \cdot 3 \cdots (n-1)}{2 \cdot 4 \cdots n} 2\pi = \frac{n!}{2^n \left[\left(\frac{n}{2} \right)! \right]^2} 2\pi$$

Por lo tanto:

$$E [X^n | X^2 + Y^2] = \begin{cases} \frac{n!}{2^n \left[\left(\frac{n}{2} \right)! \right]^2} (X^2 + Y^2)^{\frac{n}{2}} & \text{si } n \text{ es par} \\ 0 & \text{si } n \text{ es impar} \end{cases}$$

4.5. Distribuciones condicionales

Como lo mencionamos con anterioridad, al igual que toda función de densidad, la función de densidad condicional de una variable aleatoria discreta X define una distribución, la cual es llamada la distribución condicional de X dada Y .

EJEMPLO 4.25. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{N(N+1)} & \text{si } x \leq y \text{ y } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

en donde N es un entero positivo.

Encuentre la función de densidad condicional de a) X dado que $Y = y$, para $y \in \{1, \dots, N\}$, y b) Y dado que $X = x$, para $x \in \{1, \dots, N\}$.

Solución

$$a. f_Y(y) = \sum_{x=1}^N f_{X,Y}(x, y) = \sum_{x=1}^y \frac{2}{N(N+1)} = \frac{2}{N(N+1)}y$$

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{1}{y} & \text{si } x \in \{1, \dots, y\} \\ 0 & \text{en otro caso} \end{cases}$$

Así que, dado que $Y = y$, X tiene distribución uniforme en el conjunto $\{1, \dots, y\}$.

$$b. f_X(x) = \sum_{y=1}^N f_{X,Y}(x, y) = \sum_{y=x}^N \frac{2}{N(N+1)} = \frac{2}{N(N+1)}(N+1-x)$$

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \begin{cases} \frac{1}{N+1-x} & \text{si } y \in \{x, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Así que, dado que $X = x$, Y tiene distribución uniforme en el conjunto $\{x, \dots, N\}$.

EJEMPLO 4.26. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p y sea $Z = \min(X, Y)$. Para $y \in \{0, 1, \dots\}$, encuentre la función de densidad condicional de Z dado que $Y = y$.

Solución

$$\begin{aligned} f_{Z|Y}(z | y) &= \frac{P[\min(X, Y)=z, Y=y]}{P[Y=y]} = \frac{P[\min(X, y)=z, Y=y]}{P[Y=y]} P[\min(X, y) = z] \\ &= \begin{cases} P[X = z] & \text{si } z < y \\ P[X \geq z] & \text{si } z = y \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} p(1-p)^z & \text{si } z \in \{0, \dots, y-1\} \\ (1-p)^z & \text{si } z = y \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

▲

Obsérvese en el último ejemplo que si $P[Y = y] > 0$ entonces la distribución condicional de $\min(X, Y)$, dado que $Y = y$, es simplemente la distribución (no condicional) de la variable aleatoria $\min(X, y)$. Este resultado se puede generalizar. En efecto, si X y Y son dos variables aleatorias discretas independientes, y un número real tal que $P[Y = y] > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ cualquier función, entonces, para cualquier $z \in \mathbb{R}$, se tiene:

$$f_{g(X, Y)|Y}(z | y) = \frac{P[g(X, Y)=z, Y=y]}{P[Y=y]} = \frac{P[g(X, y)=z, Y=y]}{P[Y=y]} = P[g(X, y) = z]$$

Es decir, la distribución condicional de $g(X, Y)$, dado que $Y = y$, es simplemente la distribución (no condicional) de la variable aleatoria $g(X, y)$.

COMENTARIO 4.27. En general, dadas dos variables aleatorias discretas, X y Y , la distribución condicional de $g(X, Y)$, dado que $Y = y$, es también la distribución de la variable aleatoria $g(X, y)$, pero calculada tomando como función de densidad de X

a la densidad condicional $f_{X|Y}$, la cual, en general, como se muestra en los ejemplos, depende del valor de Y .

Para hacer ver la última afirmación, sean X y Y dos variables aleatorias discretas, y un número real tal que $P[Y = y] > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ cualquier función. Entonces, para cualquier $z \in \mathbb{R}$, se tiene:

$$\begin{aligned} f_{g(X,Y)|Y}(z | y) &= \frac{P[g(X,Y)=z, Y=y]}{P[Y=y]} = \frac{P[g(X,y)=z, Y=y]}{P[Y=y]} \\ &= \sum_{\{x:g(x,y)=z\}} \frac{P[X=z, Y=y]}{P[Y=y]} = \sum_{\{x:g(x,y)=z\}} f_{X|Y}(x | y). \end{aligned}$$

También se tiene:

$$\begin{aligned} F_{g(X,Y)|Y}(z | y) &= P[g(X, Y) \leq z | Y = y] = E[I_{[g(X,Y) \leq z]} | Y = y] \\ &= E[I_{\{(u,v):g(u,v) \leq z\}}(X, Y) | Y = y] = \sum_x I_{\{(u,v):g(u,v) \leq z\}}(x, y) f_{X|Y}(x | y) \\ &= \sum_x I_{\{u:g(u,y) \leq z\}}(x) f_{X|Y}(x | y) \end{aligned}$$

Es decir, $F_{g(X,Y)|Y}(z | y)$ es la función de distribución de la variable aleatoria $g(X, y)$ calculada tomando como función de densidad de X a la función de densidad condicional $f_{X|Y}$.

El comentario 4.27 muestra que la distribución condicional de una función $g(X, Y)$ de dos variables aleatorias discretas, X y Y , dado que $Y = y$, se puede tratar como una distribución no condicional si sustituimos a la función de densidad de X por la función de densidad condicional $f_{X|Y}$. Como se muestra en la siguiente proposición, esta propiedad se extiende incluso a la fórmula que da la esperanza de una variable aleatoria en términos de la integral de su función de distribución.

PROPOSICIÓN 4.28. Sean X y Y dos variables aleatorias discretas, y un número real tal que $P[Y = y] > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función tal que $g(X, Y)$ tiene esperanza finita. Entonces $\int_0^\infty [1 - F_{g(X,Y)|Y}(z | y)] dz < \infty$ y $\int_0^\infty F_{g(X,Y)|Y}(-z | y) dz < \infty$ y se tiene:

$$E[g(X, Y) | Y = y] = \int_0^\infty [1 - F_{g(X,Y)|Y}(z | y)] dz - \int_0^\infty F_{g(X,Y)|Y}(-z | y) dz$$

Demostración

Sean x_1, x_2, \dots los posibles valores de X . Por el corolario 2.43, se sabe que:

$$\begin{aligned} \sum_k |g(x_k, y)| f_{X|Y}(x_k | y) < \infty \text{ si y sólo si } \int_0^\infty [1 - F_{g(X,Y)|Y}(z | y)] dz < \infty \\ \int_0^\infty F_{g(X,Y)|Y}(-z | y) dz < \infty \end{aligned}$$

$$\sum_k g(x_k, y) f_{X|Y}(x_k | y) = \int_0^\infty [1 - F_{g(X,Y)|Y}(z | y)] dz - \int_0^\infty [1 - F_{g(X,Y)|Y}(z | y)] dz$$

Pero, como $g(X, Y)$ tiene esperanza finita, se tiene $\sum_k |g(x_k, y)| f_{X|Y}(x_k | y) < \infty$ y $E[g(X, Y) | Y = y] = \sum_k g(x_k, y) f_{X|Y}(x_k | y)$, de lo cual se obtiene el resultado. ■

EJEMPLO 4.29. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Para $x \in \{0, 1, \dots\}$, encuentre la función de densidad condicional de X dado que $X + Y = z$ e identifique la correspondiente distribución condicional.

Solución

$$\begin{aligned} f_{X|X+Y}(x | z) &= \frac{P[X=x, X+Y=z]}{P[X+Y=z]} = \frac{P[X=x]P[Y=z-x]}{P[X+Y=z]} \\ &= \begin{cases} \frac{p(1-p)^x p(1-p)^{z-x}}{(z+1)p^2(1-p)^z} & \text{si } x \in \{0, \dots, z\} \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{z+1} & \text{si } x \in \{0, \dots, z\} \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

Así que, dado que $X+Y = z$, X tiene distribución uniforme en el conjunto $\{0, \dots, z\}$. ▲

Para definir el concepto de distribución condicional en el caso absolutamente continuo se requiere de la siguiente definición:

DEFINICIÓN 4.30 (Probabilidad condicional de un evento dada una variable aleatoria). Si A es un evento cualquiera y Y cualquier variable aleatoria, se define:

$$P[A | Y] = E[I_A | Y]$$

Sea (X, Y) un vector aleatorio absolutamente continuo con función de densidad conjunta $f_{X,Y}$. Para cada $y \in \mathbb{R}$ definamos la función $x \mapsto f_{X|Y}(x | y)$ de la siguiente manera:

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ f_X(x) & \text{en otro caso} \end{cases}$$

Obsérvese que, para cualquier $y \in \mathbb{R}$, la función $x \mapsto f_{X|Y}(x | y)$ es una función de densidad. Además, si $f_Y(y) > 0$:

$$P[X \leq x | Y = y] = E[I_{(-\infty, x]}(X) | Y = y]$$

$$= \int_{-\infty}^{\infty} I_{(-\infty, x]}(u) f_{X|Y}(u | y) du = \int_{-\infty}^x f_{X|Y}(u | y) du.$$

Por lo tanto, al igual que en el caso discreto, resulta natural definir a la función $x \mapsto f_{X|Y}(x | y)$ como una **función de densidad condicional de X dado que Y = y** y a la distribución que define como la **distribución condicional de X dada Y**. También definimos a la función $x \mapsto F_{X|Y}(x | y) = P[X \leq x | Y = y]$ como la **función de distribución condicional de X dado que Y = y**.

COMENTARIO 4.31. *Obsérvese que, al igual que en el caso discreto, la definición de una función de densidad condicional, en el caso absolutamente continuo, está acorde con la definición de la esperanza condicional. En efecto, si y es un número real tal que $f_Y(y) > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función tal que $g(X, Y)$ tiene esperanza finita y $\int_{-\infty}^{\infty} |g(x, y)| f_{X|Y}(x | y) dx < \infty$, entonces se tiene:*

$$E[g(X, Y) | Y = y] = \int_{-\infty}^{\infty} g(x, y) \frac{f_{X,Y}(x, y)}{f_Y(y)} dx = \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x | y) dx$$

Es decir, dado que $Y = y$, la esperanza condicional de $g(X, Y)$ es la esperanza de la variable aleatoria $g(X, y)$ calculada tomando como función de densidad de X a la densidad condicional $f_{X|Y}$.

EJEMPLO 4.32. *Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:*

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda x} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) la distribución condicional de X dado que $Y = y$, para $y > 0$, y b) la distribución condicional de Y dado que $X = x$, para $x > 0$.

Solución

$$a. f_Y(y) = \begin{cases} \int_y^{\infty} \lambda^2 e^{-\lambda x} dx & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \lambda e^{-\lambda y} & \text{si } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{X|Y}(x | y) = \begin{cases} \frac{f_{X,Y}(x, y)}{f_Y(y)} & \text{si } f_Y(y) > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{\lambda^2 e^{-\lambda x}}{\lambda e^{-\lambda y}} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \lambda e^{-\lambda(x-y)} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, dado que $Y = y$, $X - y$ tiene distribución exponencial de parámetro λ .

$$b. f_X(x) = \begin{cases} \int_0^x \lambda^2 e^{-\lambda x} dy & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \lambda^2 x e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{Y|X}(y | x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{si } f_X(x) > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{\lambda^2 e^{-\lambda x}}{\lambda^2 x e^{-\lambda x}} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{1}{x} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, dado que $X = x$, Y tiene distribución uniforme en el intervalo $(0, x)$.

EJEMPLO 4.33. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas (μ_X, μ_Y) , vector de varianzas (σ_X^2, σ_Y^2) y coeficiente de correlación ρ . Demuestre que la distribución condicional de Y , dado que $X = x$, es normal con media $\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ y varianza $\sigma_Y^2(1 - \rho^2)$.

Solución

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \sqrt{2\pi}\sigma_X \exp \left[\frac{1}{2\sigma_X^2}(x - \mu_X)^2 \right] \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

$$\exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ \frac{1}{2\sigma_X^2}(x - \mu_X)^2 - \frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\rho^2 \left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2\sigma_Y^2(1-\rho^2)} \left[\rho^2\sigma_Y^2 \left(\frac{x-\mu_X}{\sigma_X} \right)^2 + (y - \mu_Y)^2 - 2\rho\sigma_Y \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X} \right] \right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2\sigma_Y^2(1-\rho^2)} \left[(y - \mu_Y) - \rho\sigma_Y \left(\frac{x-\mu_X}{\sigma_X} \right) \right]^2 \right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2\sigma_Y^2(1-\rho^2)} \left[y - \left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) \right) \right]^2 \right\}$$

EJEMPLO 4.34. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 1)$ y coeficiente de correlación $\frac{1}{2}$. Encuentre e identifique la distribución de $2X + Y$ dado que $2Y - X = z$, para cualquier $z \in \mathbb{R}$.

Solución

El vector $(2X + Y, 2Y - X)$ se obtiene del vector (X, Y) mediante una transformación lineal invertible. Por lo tanto, la distribución de $(2X + Y, 2Y - X)$ es también normal bivariada. Dado que $2Y - X = z$, $2X + Y$ tiene entonces distribución normal con media $\mu_{2X+Y} + \rho \frac{\sigma_{2X+Y}}{\sigma_{2Y-X}} (z - \mu_{2Y-X})$ y varianza $\sigma_{2X+Y}^2 (1 - \rho^2)$, en donde ρ es el coeficiente de correlación de la pareja $(2X + Y, 2Y - X)$. Además, se tiene:

$$\mu_{2X+Y} = \mu_{2Y-X} = 0$$

$$\sigma_{2X+Y}^2 = 4\sigma_X^2 + \sigma_Y^2 + 4Cov(X, Y) = 7$$

$$\sigma_{2Y-X}^2 = \sigma_X^2 + 4\sigma_Y^2 - 4Cov(X, Y) = 3$$

$$\begin{aligned} Cov(2X + Y, 2Y - X) &= E[(2X + Y)(2Y - X)] \\ &= 3E[XY] - 2E[X^2] + 2E[Y^2] = 3Cov(X, Y) - 2\sigma_X^2 + 2\sigma_Y^2 = \frac{3}{2} \end{aligned}$$

$$\rho = \frac{Cov(2X+Y, 2Y-X)}{\sigma_{2X+Y}\sigma_{2Y-X}} = \frac{1}{14}\sqrt{21}$$

Por lo tanto, dado que $2Y - X = z$, $2X + Y$ tiene distribución normal con media $\frac{1}{2}z$ y varianza $\frac{25}{4}$.

EJEMPLO 4.35. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama, X de parámetros α y λ , Y de parámetros β y λ . Encuentre la distribución condicional de X dado que $X + Y = z$, para $z > 0$. ¿Cuál es el mejor estimador de X , en el sentido de la media cuadrática, dado que $X + Y = z$?

Solución

$$\begin{aligned} f_{X, X+Y}(x, z) &= f_X(x)f_Y(z-x) \\ &= \begin{cases} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(z-x)^{\beta-1} e^{-\lambda z} & \text{si } 0 < x < z \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

$$f_{X+Y}(z) = \begin{cases} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\lambda z} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{X|X+Y}(x | z) = \begin{cases} \frac{1}{B(\alpha, \beta)} z^{1-\alpha-\beta} x^{\alpha-1} (z-x)^{\beta-1} & \text{si } 0 < x < z \\ 0 & \text{en otro caso} \end{cases}$$

Observemos que si z es una constante positiva y U es una variable aleatoria con distribución beta de parámetros α y β , entonces una función de densidad de la variable aleatoria zU está dada por:

$$f_{zU}(x) = \begin{cases} \frac{1}{z} \frac{1}{B(\alpha, \beta)} \left(\frac{x}{z}\right)^{\alpha-1} \left(1 - \frac{x}{z}\right)^{\beta-1} & \text{si } 0 < x < z \\ 0 & \text{en otro caso} \end{cases}$$

$$= \begin{cases} \frac{1}{B(\alpha, \beta)} z^{1-\alpha-\beta} x^{\alpha-1} (z-x)^{\beta-1} & \text{si } 0 < x < z \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, dado que $X + Y = z$, $\frac{X}{z}$ tiene distribución beta con parámetros α y β .

Con base en este resultado, se tiene:

$$E\left[\frac{X}{z} \mid X + Y = z\right] = \frac{\alpha}{\alpha + \beta}.$$

$$\text{Así que, } E[X \mid X + Y = z] = \frac{\alpha z}{\alpha + \beta}.$$

De manera que, dado que $X + Y = z$, el mejor estimador de X , en el sentido de la media cuadrática, es $\frac{\alpha z}{\alpha + \beta}$.

EJEMPLO 4.36. Sean X y Y variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre la distribución condicional de X dado que $Y - X = z$, para cualquier $z \in \mathbb{R}$.

Solución

$$f_{X, Y-X}(x, z) = f_X(x) f_Y(z+x) = \begin{cases} \lambda^2 e^{-\lambda z} e^{-2\lambda x} & \text{si } x > \max(-z, 0) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{Y-X}(z) = \begin{cases} \int_0^\infty \lambda^2 e^{-\lambda z} e^{-2\lambda x} dx & \text{si } z > 0 \\ \int_{-z}^\infty \lambda^2 e^{-\lambda z} e^{-2\lambda x} dx & \text{si } z \leq 0 \end{cases} = \begin{cases} \frac{1}{2} \lambda e^{-\lambda z} & \text{si } z > 0 \\ \frac{1}{2} \lambda e^{\lambda z} & \text{si } z \leq 0 \end{cases}$$

$$f_{X|Y-X}(x \mid z) = \begin{cases} 2\lambda e^{-2\lambda x} & \text{si } z > 0, x > 0 \\ 2\lambda e^{-2\lambda(x+z)} & \text{si } z \leq 0, x > -z \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, dado que $Y - X = z$, $X + z$ tiene distribución exponencial con parámetro 2λ si $z \leq 0$ y X tiene distribución exponencial con parámetro 2λ si $z > 0$.

EJEMPLO 4.37. Sean X y Y variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre a) una función de densidad condicional de $X + Y$ dado que $Y - X = v$, para cualquier $v \in \mathbb{R}$, y b) $E[(X + Y)^2 \mid Y - X]$.

Solución

$$a. f_{X+Y, Y-X}(u, v) = \frac{1}{2} f_X\left(\frac{u-v}{2}\right) f_Y\left(\frac{u+v}{2}\right)$$

$$= \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda u} & \text{si } -u < v < u, u > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{Y-X}(v) = \int_0^\infty f_{X+Y, Y-X}(u, v) du = \frac{1}{2} \int_{|v|}^\infty \lambda^2 e^{-\lambda u} du = \frac{1}{2} \lambda e^{-\lambda|v|}$$

$$f_{X+Y|Y-X}(u | v) = \frac{f_{X+Y, Y-X}(u, v)}{f_{Y-X}(v)} = \begin{cases} \lambda e^{-\lambda(u-|v|)} & \text{si } u > |v| \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto, dado que $Y - X = v$, $X + Y - |v|$ tiene distribución exponencial con parámetro λ .

b. Dado $Y - X = v$, sea $Z = X + Y - |v|$, entonces:

$$\begin{aligned} E[(X + Y)^2 | Y - X = v] &= E[(Z + |v|)^2 | Y - X = v] \\ &= E[Z^2 | Y - X = v] + 2|v| E[Z | Y - X = v] + v^2 \\ &= \frac{2}{\lambda^2} + \frac{2}{\lambda} |v| + v^2 \end{aligned}$$

$$\text{Por lo tanto, } E[(X + Y)^2 | Y - X] = \frac{2}{\lambda^2} + \frac{2}{\lambda} |Y - X| + (Y - X)^2.$$

EJEMPLO 4.38. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$ y sea $Z = \min(X, Y)$. Para $y \in (0, 1)$, encuentre la función de distribución condicional de Z dado que $Y = y$.

Solución

Obsérvese que la función de distribución conjunta de Z y Y no es absolutamente continua. En efecto, si lo fuera, se tendría $P[Z = Y] = 0$, pero $P[Z = Y] = P[X \geq Y] = \frac{1}{2}$. La distribución de Z , dado el valor de Y , no puede entonces obtenerse mediante la función de densidad condicional $f_{Z|Y}$.

Se tiene:

$$I_{[\min(X, Y) > z]} = I_{[X > z]} I_{[Y > z]} = I_{(z, \infty)}(X) I_{(z, \infty)}(Y)$$

Así que:

$$\begin{aligned} P[\min(X, Y) > z | Y = y] &= E[I_{[\min(X, Y) > z]} | Y = y] \\ &= E[I_{(z, \infty)}(X) I_{(z, \infty)}(Y) | Y = y] = \int_{-\infty}^\infty I_{(z, \infty)}(x) I_{(z, \infty)}(y) f_{X|Y}(x | y) dx \\ &= \int_0^1 I_{(z, \infty)}(x) I_{(z, \infty)}(y) \frac{f_{X, Y}(x, y)}{f_Y(y)} dx = \int_0^1 I_{(z, \infty)}(x) I_{(z, \infty)}(y) f_X(x) dx \end{aligned}$$

$$= \int_0^1 I_{(z,1)}(x)I_{(z,1)}(y)dx = \begin{cases} 1 & \text{si } z \leq 0 \\ 1 - z & \text{si } 0 < z < y \\ 0 & \text{si } z \geq y \end{cases}$$

Por lo tanto:

$$F_{Z|Y}(z | y) = P[\text{mín}(X, Y) \leq z | Y = y] = \begin{cases} 0 & \text{si } z \leq 0 \\ z & \text{si } 0 < z < y \\ 1 & \text{si } z \geq y \end{cases}$$

Obsérvese que, dado $Y = y$, la función de distribución condicional de Z no es ni discreta ni continua. En efecto, evidentemente no es discreta y tiene una discontinuidad en $z = y$.

▲

Obsérvese en el último ejemplo que si $f_Y(y) > 0$ entonces la distribución condicional de $\text{mín}(X, Y)$, dado que $Y = y$, es simplemente la distribución (no condicional) de la variable aleatoria $\text{mín}(X, y)$. Este resultado se puede generalizar. En efecto, si X y Y son dos variables aleatorias absolutamente continuas independientes, y un número real tal que $f_Y(y) > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ cualquier función, entonces, para cualquier $z \in \mathbb{R}$, se tiene:

$$\begin{aligned} F_{g(X,Y)|Y}(z | y) &= P[g(X, Y) \leq z | Y = y] = E[I_{[g(X,Y) \leq z]} | Y = y] \\ &= E[I_{[g(X,y) \leq z]}] = P[g(X, y) \leq z] = F_{g(X,y)}(z) \end{aligned}$$

Es decir, la distribución condicional de $g(X, Y)$, dado que $Y = y$, es simplemente la distribución (no condicional) de la variable aleatoria $g(X, y)$.

COMENTARIO 4.39. En general, dado un vector aleatorio absolutamente continuo, (X, Y) , la distribución condicional de $g(X, Y)$, dado que $Y = y$, es también la distribución de la variable aleatoria $g(X, y)$, pero calculada tomando como función de densidad de X a la función de densidad condicional $f_{X|Y}$, la cual, en general, como se muestra en los ejemplos, depende del valor de Y .

Para hacer ver la última afirmación, sea (X, Y) un vector aleatorio absolutamente continuo, y un número real tal que $P[Y = y] > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ cualquier función. Entonces, para cualquier $z \in \mathbb{R}$, se tiene:

$$\begin{aligned} F_{g(X,Y)|Y}(z | y) &= P[g(X, Y) \leq z | Y = y] = E[I_{\{g(X,Y) \leq z\}} | Y = y] \\ &= E[I_{\{(u,v):g(u,v) \leq z\}}(X, Y) | Y = y] \\ &= \int_{-\infty}^{\infty} I_{\{(u,v):g(u,v) \leq z\}}(x, y) f_{X|Y}(x | y) dx \\ &= \int_{-\infty}^{\infty} I_{\{u:g(u,y) \leq z\}}(x) f_{X|Y}(x | y) dx \end{aligned}$$

Es decir, $F_{g(X,Y)|Y}(z | y)$ es la función de distribución de la variable aleatoria $g(X, y)$, pero calculada tomando como función de densidad de X a la función de densidad condicional $f_{X|Y}$.

Los comentarios 4.31 y 4.39 muestran que, al igual que en el caso discreto, la distribución condicional de una función $g(X, Y)$ de un vector aleatorio absolutamente continuo, (X, Y) , dado que $Y = y$, se puede tratar como una distribución no condicional si sustituimos a la función de densidad de X por la función de densidad condicional $f_{X|Y}$. Como se muestra en la siguiente proposición, al igual que en el caso discreto, esta propiedad se extiende incluso a la fórmula que da la esperanza de una variable aleatoria en términos de la integral de su función de distribución.

PROPOSICIÓN 4.40. Sea (X, Y) un vector aleatorio absolutamente continuo, y un número real tal que $f_Y(y) > 0$ y $g : \mathbb{R}^2 \mapsto \mathbb{R}$ una función tal que $g(X, Y)$ tiene esperanza finita y $\int_{-\infty}^{\infty} |g(x, y)| f_{X|Y}(x | y) dx < \infty$. Entonces:

$$\begin{aligned} \int_0^{\infty} [1 - F_{g(X,Y)|Y}(z | y)] dz &< \infty \\ \int_0^{\infty} F_{g(X,Y)|Y}(-z | y) dz &< \infty \\ E[g(X, Y) | Y = y] &= \int_0^{\infty} [1 - F_{g(X,Y)|Y}(z | y)] dz - \int_0^{\infty} F_{g(X,Y)|Y}(-z | y) dz \end{aligned}$$

Demostración

Se sabe, por el corolario 2.45, que $\int_{-\infty}^{\infty} |g(x, y)| f_{X|Y}(x | y) dx < \infty$ si y sólo si $\int_0^{\infty} [1 - F_{g(X,Y)|Y}(z | y)] dz < \infty$ y $\int_0^{\infty} F_{g(X,Y)|Y}(-z | y) dz < \infty$ y, en ese caso:

$$\int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x | y) dx = \int_0^{\infty} [1 - F_{g(X,Y)|Y}(z | y)] dz - \int_0^{\infty} [1 - F_{g(X,Y)|Y}(z | y)] dz$$

Pero se tiene $E[g(X, Y) | Y = y] = \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x | y) dx$, de lo cual se sigue el resultado. ■

En el caso del ejemplo 4.38, se tiene:

$$E[Z | Y = y] = \int_0^y [1 - F_{Z|Y}(z | y)] dz = \int_0^y (1 - z) dz = y - \frac{1}{2}y^2$$

Por lo tanto, $E[Z | Y] = Y - \frac{1}{2}Y^2$.

EJEMPLO 4.41. *Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t , entonces la familia de variables aleatorias $\{X_t\}_{t \geq 0}$ forma un proceso de Poisson de parámetro λ . Vamos a encontrar la distribución conjunta de T_1, \dots, T_n , dado que $T_{n+1} = t$, en donde $t > 0$ y $n \in N$.*

Recordemos que:

$$f_{T_1, \dots, T_{n+1}}(t_1, \dots, t_{n+1}) = \begin{cases} \lambda^n e^{-\lambda t_n} & \text{si } 0 < t_1 < \dots < t_{n+1} \\ 0 & \text{en otro caso} \end{cases}$$

Además, $T_{n+1} = T_1 + (T_2 - T_1) + \dots + (T_{n+1} - T_n)$, así que T_{n+1} tiene distribución gama con parámetros $\alpha = n + 1$ y λ .

Sean $0 < t_1 < t_2 < \dots < t_n$, entonces:

$$\begin{aligned} f_{T_1, \dots, T_n | T_{n+1}}(t_1, \dots, t_n | t) &= \frac{f_{T_1, \dots, T_{n+1}}(t_1, \dots, t_n, t)}{f_{T_{n+1}}(t)} \\ &= \begin{cases} \frac{n! \lambda^{n+1} e^{-\lambda t}}{\lambda^{n+1} t^n e^{-\lambda t}} & \text{si } 0 < t_1 < \dots < t_n < t \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{n!}{t^n} & \text{si } 0 < t_1 < \dots < t_n < t \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

Por lo tanto, dado que $T_{n+1} = t$, la distribución conjunta de T_1, \dots, T_n es la misma que la de los estadísticos de orden correspondientes a n variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, t)$.

4.6. Regla general de la probabilidad total

Sea X una variable aleatoria de esperanza finita y Y cualquier variable aleatoria. Entonces sabemos que la esperanza condicional $E[X | Y]$ existe y, por el inciso *iv* de la proposición 4.14, se tiene:

$$E[X] = E[E[X | Y]]$$

Esta propiedad de la esperanza condicional resulta sumamente útil en la solución de muchos problemas.

Obsérvese que, en el caso en que la variable aleatoria Y sea discreta, tal propiedad se expresa de la siguiente manera:

$$E[X] = \sum_y E[X | Y = y] P[Y = y]$$

Fórmula que generaliza la regla de la probabilidad total.

En el caso en que Y sea una variable aleatoria absolutamente continua, se tiene:

$$E[X] = \int_{-\infty}^{\infty} E[X | Y = y] f_Y(y) dy$$

Fórmula que también puede verse como una generalización, al caso continuo, de la regla de la probabilidad total.

Con base en lo anterior, la relación $E[X] = E[E[X | Y]]$ será llamada en lo sucesivo la **regla general de la probabilidad total**.

EJEMPLO 4.42. *Sea N una variable aleatoria con distribución binomial de parámetros m y q y supongamos que, para cada valor n de N , X es una variable aleatoria con distribución binomial de parámetros n y p . Encuentre $E[X]$ y $Var(X)$.*

Solución

$$\begin{aligned} E[X] &= E[E[X | N]] = \sum_{n=0}^m E[X | N = n] P[N = n] \\ &= \sum_{n=0}^m npP[N = n] = pE[N] = mpq \\ E[X^2] &= E[E[X^2 | N]] = \sum_{n=0}^m E[X^2 | N = n] P[N = n] \\ &= \sum_{n=0}^m [np(1-p) + n^2p^2] P[N = n] \\ &= p(1-p)E[N] + p^2E[N^2] = p(1-p)mq + p^2[mq(1-q) + m^2q^2] \end{aligned}$$

Así que:

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - [E(X)]^2 \\
 &= p(1-p)mq + p^2[mq(1-q) + m^2q^2] - m^2p^2q^2 \\
 &= p(1-p)mq + p^2mq(1-q) = mpq(1-pq)
 \end{aligned}$$

EJEMPLO 4.43. Sea Y una variables aleatoria con distribución exponencial de parámetro λ y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución exponencial de parámetro $\frac{1}{y}$. Encuentre $E[X]$ y $\text{Var}(X)$.

Solución

$$\begin{aligned}
 E[X] &= E[E[X | Y]] = \int_0^\infty E[X | Y = y] f_Y(y) dy \\
 &= \int_0^\infty y f_Y(y) dy = E[Y] = \frac{1}{\lambda}. \\
 E[X^2] &= E[E[X^2 | Y]] = \int_0^\infty E[X^2 | Y = y] f_Y(y) dy \\
 &= \int_0^\infty 2y^2 f_Y(y) dy = 2E[Y^2] = \frac{4}{\lambda^2}
 \end{aligned}$$

Así que:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{4}{\lambda^2} - \frac{1}{\lambda^2} = \frac{3}{\lambda^2}$$

EJEMPLO 4.44. Una urna contiene 10 bolas rojas y 20 bolas negras. Se van seleccionando bolas de la urna al azar, una a una y con reemplazo, hasta que se obtienen 4 bolas rojas en forma consecutiva. Si X es el número de bolas seleccionadas hasta que se detiene el proceso, encuentre $E[X]$.

Solución

Sea Y el número de elecciones que se realizan hasta obtener una bola negra por primera vez. entonces:

$$P[Y = k] = \begin{cases} \frac{2}{3} \left(\frac{1}{3}\right)^{k-1} & \text{si } k \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

Considerando ahora que si la primera bola negra se obtiene en alguna de las primeras 4 elecciones entonces se vuelve a la situación del inicio, se tiene, para $k \in \mathbb{N}$:

$$E[X | Y = k] = \begin{cases} k + E[X] & \text{si } k \in \{1, 2, 3, 4\} \\ 4 & \text{en otro caso} \end{cases}$$

Así que:

$$E[X] = \sum_{k=1}^{\infty} E[X | Y = k] P[Y = k]$$

$$\begin{aligned}
&= \sum_{k=1}^4 (k + E[X]) \frac{2}{3} \left(\frac{1}{3}\right)^{k-1} + \sum_{k=5}^{\infty} 4 \frac{2}{3} \left(\frac{1}{3}\right)^{k-1} \\
&= \frac{2}{3} \sum_{k=1}^4 k \left(\frac{1}{3}\right)^{k-1} + \frac{2}{3} E[X] \sum_{k=1}^4 \left(\frac{1}{3}\right)^{k-1} + \frac{8}{3} \sum_{k=5}^{\infty} \left(\frac{1}{3}\right)^{k-1} \\
&= \frac{116}{81} + \frac{80}{81} E[X] + \frac{4}{81} = \frac{40}{27} + \frac{80}{81} E[X]
\end{aligned}$$

Por lo tanto, $E[X] = 120$.

EJEMPLO 4.45. Una urna contiene inicialmente a bolas azules y r bolas rojas. Se agregan s bolas rojas a la urna e inmediatamente después se seleccionan, al azar y sin reemplazo, s bolas de la misma. Supongamos que este proceso se repite indefinidamente y llamemos X_n al número de bolas azules que quedan en la urna después del paso n . Encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.

Solución

Después del paso n , hay $a+r$ bolas en la urna, de las cuales X_n son azules. Al agregar s bolas rojas, quedan en la urna $a+r+s$ bolas, de las cuales X_n son azules. Al tomar una muestra sin reemplazo de s bolas de esta urna, la distribución del número de bolas azules que salen en la muestra es hipergeométrica, de manera que su valor esperado está dado por $\frac{sX_n}{a+r+s}$. Así que:

$$E[X_{n+1} | X_n] = X_n - \frac{sX_n}{a+r+s} = \frac{a+r}{a+r+s} X_n$$

Por lo tanto, para $n \in \mathbb{N}$, se tiene:

$$E[X_{n+1}] = E[E(X_{n+1} | X_n)] = \frac{a+r}{a+r+s} E[X_n]$$

Así que:

$$E[X_n] = \left(\frac{a+r}{a+r+s}\right)^{n-1} E[X_1] = \left(\frac{a+r}{a+r+s}\right)^{n-1} \frac{a+r}{a+r+s} a \left(\frac{a+r}{a+r+s}\right)^n$$

EJEMPLO 4.46. Supongamos que el número de personas que entran a un elevador, en la planta baja de un edificio de N pisos, tiene distribución Poisson con parámetro λ . Supongamos, además, que cada persona que sube al elevador baja de él, al azar, en cualquiera de los N pisos, independientemente de donde bajen las otras personas. Encuentre el número esperado de paradas que hace el elevador hasta que bajan todas las personas.

Solución

Sea Y el número de personas que suben al elevador en la planta baja y X el número de paradas que hace el elevador hasta que bajan todas las personas y definamos las variables aleatorias X_1, X_2, \dots, X_N de la siguiente manera:

$$X_i = \begin{cases} 1 & \text{si el elevador para en el piso } i \\ 0 & \text{si el elevador no para en el piso } i \end{cases}$$

Se tiene $X = \sum_{i=1}^N X_i$ y, para $k \in \{0, 1, \dots\}$:

$$E[X_i | Y = k] = 1 - \left(1 - \frac{1}{N}\right)^k$$

Así que:

$$E[X | Y = k] = \sum_{i=1}^N E[X_i | Y = k] = N \left[1 - \left(1 - \frac{1}{N}\right)^k\right]$$

Por lo tanto:

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} E[X | Y = k] P[Y = k] = \sum_{k=0}^{\infty} N \left[1 - \left(1 - \frac{1}{N}\right)^k\right] \frac{e^{-\lambda} \lambda^k}{k!} \\ &= Ne^{-\lambda} \left[\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \sum_{k=0}^{\infty} \frac{\lambda^k \left(1 - \frac{1}{N}\right)^k}{k!} \right] = Ne^{-\lambda} \left[e^{\lambda} - e^{\lambda \left(1 - \frac{1}{N}\right)} \right] \\ &= N \left[1 - e^{-\frac{\lambda}{N}} \right] \end{aligned}$$

EJEMPLO 4.47. Sea Y_1 un número que se elige al azar en el intervalo $(0, 1)$, Y_2 un número que se elige al azar en el intervalo $(1 - Y_1, 1)$, Y_3 un número que se elige al azar en el intervalo $(1 - Y_2, 1)$, etc. Encuentre $E[Y_n]$ para cualquier $n \in \mathbb{N}$.

Solución

Para $n \in \mathbb{N}$, se tiene:

$$E[Y_{n+1}] = E[E(Y_{n+1} | Y_n)] = E\left[1 - \frac{1}{2}Y_n\right] = 1 - \frac{1}{2}E[Y_n]$$

Por lo tanto:

$$E[Y_n] = \sum_{k=0}^n (-1)^k \frac{1}{2^k} = \frac{2}{3} \left[1 - (-1)^{n+1} \frac{1}{2^{n+1}}\right]$$

▲

En el caso en que el vector aleatorio (X, Y) sea discreto, la regla de la probabilidad total nos permite escribir la relación:

$$f_X(x) = \sum_y f_{X|Y}(x | y) f_Y(y)$$

para cualquier $x \in \mathbb{R}$

En el caso absolutamente continuo se tiene la relación análoga, de acuerdo con la siguiente proposición:

PROPOSICIÓN 4.48. Sea (X, Y) un vector aleatorio absolutamente continuo, entonces, para cualquier $x \in \mathbb{R}$, la función:

$$x \mapsto \int_{-\infty}^{\infty} f_{X|Y}(x | y) f_Y(y) dy$$

es una función de densidad de X .

Demostración

$$\begin{aligned} P[X \leq x] &= E[I_{[X \leq x]}] = \int_{-\infty}^{\infty} E[I_{[X \leq x]} | Y = y] f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P[X \leq x | Y = y] f_Y(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^x f_{X|Y}(u | y) f_Y(y) du dy \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_{X|Y}(u | y) f_Y(y) dy du \end{aligned}$$

de lo cual se sigue el resultado. ■

EJEMPLO 4.49. Sea N una variable aleatoria con distribución binomial de parámetros m y q y supongamos que, para cada valor n de N , X es una variable aleatoria con distribución binomial de parámetros n y p . Encuentre la función de densidad de X .

Solución

Para $k \in \{0, \dots, m\}$, se tiene:

$$\begin{aligned} P[X = k] &= \sum_{n=k}^m P[X = k | N = n] P[N = n] \\ &= \sum_{n=k}^m \binom{n}{k} p^k (1-p)^{n-k} \binom{m}{n} q^n (1-q)^{m-n} \\ &= \binom{m}{k} (pq)^k \sum_{n=k}^m \frac{\binom{n}{k} \binom{m}{n}}{\binom{m}{k}} [q(1-p)]^{n-k} (1-q)^{m-n} \\ &= \binom{m}{k} (pq)^k \sum_{n=0}^{m-k} \frac{\binom{n+k}{k} \binom{m}{n+k}}{\binom{m}{k}} [q(1-p)]^n (1-q)^{m-k-n} \\ &= \binom{m}{k} (pq)^k \sum_{n=0}^{m-k} \binom{m-k}{n} [q(1-p)]^n (1-q)^{m-k-n} \\ &= \binom{m}{k} (pq)^k [q(1-p) + (1-q)]^{m-k} = \binom{m}{k} (pq)^k [1-pq]^{m-k} \end{aligned}$$

Así que, X tiene distribución binomial con parámetros m y pq .

El resultado puede interpretarse de la siguiente manera:

Para evaluar N , se realizan m ensayos de Bernoulli independientes con probabilidad de éxito q en cada ensayo. El valor de X puede entonces obtenerse de la siguiente manera: Comenzando con el valor $X = 0$, para cada uno de los m ensayos, si hay

éxito, se realiza un ensayo de Bernoulli, independiente de cualquier otro ensayo, con probabilidad de éxito p . Si hay éxito en este último ensayo, entonces el valor de X se incrementa en 1. En otras palabras, el valor de X se incrementa en 1 únicamente cuando hay éxito en ambos ensayos de Bernoulli, lo cual ocurre con probabilidad pq . Así, X cuenta el número de éxitos en una sucesión de m ensayos de Bernoulli independientes, en cada uno de los cuales la probabilidad de éxito es igual a pq .

EJEMPLO 4.50. Sea Y una variable aleatoria con distribución exponencial de parámetro λ y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución exponencial de parámetro y . Encuentre a) una función de densidad de X , b) una función de densidad condicional de Y dado que $X = x$, para cada $x \in \mathbb{R}$ tal que $f_X(x) > 0$, y c) $E[Y | X]$.

Solución

a. Se tiene:

$$f_{X|Y}(x | y) = \begin{cases} ye^{-xy} & \text{si } x > 0 \text{ y } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que:

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|Y}(x | y)f_Y(y)dy = \begin{cases} \int_0^\infty ye^{-xy}\lambda e^{-\lambda y}dy & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \int_0^\infty y\lambda e^{-(\lambda+x)y}dy & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} = \begin{cases} \frac{\lambda}{(\lambda+x)^2} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

b. Se tiene:

$$f_{X,Y}(x, y) = f_{X|Y}(x | y)f_Y(y) = \begin{cases} \lambda ye^{-y(\lambda+x)} & \text{si } x > 0 \text{ y } y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Así que:

$$\begin{aligned} f_{Y|X}(y | x) &= \begin{cases} \frac{\lambda ye^{-y(\lambda+x)}}{\frac{\lambda}{(\lambda+x)^2}} & \text{si } x > 0 \text{ y } y > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} (\lambda+x)^2 ye^{-y(\lambda+x)} & \text{si } x > 0 \text{ y } y > 0 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

c. De acuerdo con el resultado en la parte b, si $x \in \mathbb{R}$ es tal que $f_X(x) > 0$, entonces, dado que $X = x$, Y tiene distribución gama con parámetros 2 y $\lambda + x$. Así que, $E[Y | X] = \frac{2}{\lambda + X}$.

EJEMPLO 4.51. Sea X un número que se elige al azar en el intervalo $(0, 1)$ y Y un número que se elige al azar en el intervalo $(1 - X, 1)$. Encuentre una función de densidad de Y .

Solución

Se tiene:

$$f_{Y|X}(y | x) = \begin{cases} \frac{1}{x} & \text{si } 0 < 1 - x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Por lo tanto:

$$\begin{aligned} f_Y(y) &= \int_0^1 f_{Y|X}(y | x) f_X(x) dx = \begin{cases} \int_{1-y}^1 \frac{1}{x} dx & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} -\ln(1 - y) & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

EJEMPLO 4.52 (Procesos de ramificación). Supongamos que una persona tiene un gen, el cual se produjo por una mutación de un gen que se transmite de generación en generación. Esta persona tiene descendientes, cada uno de los cuales puede o no poseer el gen mutado. Supongamos además que, considerando únicamente las personas de la población que poseen el gen mutado en un momento dado, el número de hijos, que poseen el gen mutado, de cada individuo de la población es independiente del número de personas en la población y del número de hijos, que poseen el gen mutado, de los otros individuos de la población y está dado por una variable aleatoria Z de esperanza finita tal que $P[Z = 0] > 0$ y $P[Z = 0] + P[Z = 1] < 1$. Consideremos a los descendientes de la persona en consideración por generaciones, siendo los hijos la primera generación, los nietos la segunda, etc. Para cada $n \in \mathbb{N}$, sea X_n el número de individuos en la generación n que poseen el gen mutado. Vamos a encontrar $E[X_n]$ y $P[\lim_{n \rightarrow \infty} X_n = 0]$. Este último límite es llamado la probabilidad de extinción del gen mutado.

Para $k \in \mathbb{N}$, dado que $X_{n-1} = k$, X_n se puede expresar como la suma de k variables aleatorias independientes Z_1, \dots, Z_k cada una de las cuales tiene la misma distribución que Z y que también son independientes de X_{n-1} , así que:

$$\begin{aligned} E[X_n] &= \sum_{k=1}^{\infty} E[X_n | X_{n-1} = k] P[X_{n-1} = k] \\ &= \sum_{k=1}^{\infty} E\left[\sum_{j=1}^k Z_j | X_{n-1} = k\right] P[X_{n-1} = k] \\ &= \sum_{k=1}^{\infty} \left(\sum_{j=1}^k E[Z_j]\right) P[X_{n-1} = k] \end{aligned}$$

$$= \sum_{k=1}^{\infty} (kE[Z]) P[X_{n-1} = k] = E[Z] \sum_{k=1}^{\infty} kP[X_{n-1} = k] = E[Z] E[X_{n-1}]$$

Por lo tanto, $E[X_n] = (E[Z])^n$.

Para la segunda parte, obsérvese que:

$$\begin{aligned} & P[\lim_{n \rightarrow \infty} X_n = 0] \\ &= P[\{\omega \in \Omega : \text{Existe } N(\omega) \text{ tal que } X_n(\omega) = 0 \text{ para cualquier } n \geq N(\omega)\}] \\ &= P[\{\omega \in \Omega : X_n(\omega) = 0 \text{ para alguna } n \in \mathbb{N}\}] \\ &= P(\bigcup_{n=1}^{\infty} [X_n = 0]) \end{aligned}$$

Además la sucesión de eventos $[X_n = 0]$ es monótona creciente, así que:

$$P(\bigcup_{n=1}^{\infty} [X_n = 0]) = \lim_{n \rightarrow \infty} P[X_n = 0]$$

Por lo tanto:

$$P[\lim_{n \rightarrow \infty} X_n = 0] = \lim_{n \rightarrow \infty} P[X_n = 0]$$

Sea Φ la función generadora de probabilidades de Z . Entonces Φ es continua en el intervalo $[-1, 1]$ y diferenciable en el intervalo $(-1, 1)$. Además, como $P[Z = 0] + P[Z = 1] < 1$, entonces $\Phi'(t) > 0$ para cualquier $t \in (0, 1)$, así que Φ es estrictamente creciente en el intervalo $[0, 1]$.

Por otra parte:

$$\begin{aligned} E[t^{X_n} | X_{n-1} = k] &= E[t^{\sum_{j=1}^k Z_j} | X_{n-1} = k] = E[t^{\sum_{j=1}^k Z_j}] \\ &= \Phi_{Z_1}(t)\Phi_{Z_2}(t) \cdots \Phi_{Z_k}(t) = [\Phi(t)]^k \end{aligned}$$

Así que:

$$\Phi_{X_n}(t) = E[t^{X_n}] = E(E[t^{X_n} | X_{n-1}]) = E([\Phi(t)]^{X_{n-1}}) = \Phi_{X_{n-1}}(\Phi(t))$$

Por lo tanto:

$$\begin{aligned} \Phi_{X_n}(t) &= (\Phi_{X_{n-1}} \circ \Phi)(t) = (\Phi_{X_{n-2}} \circ \Phi \circ \Phi)(t) \\ &= \cdots = (\Phi_{X_1} \circ \Phi^{(n-1)})(t) = \Phi^{(n)}(t) \end{aligned}$$

Sea $p_n = P[X_n = 0] = \Phi_{X_n}(0) = \Phi^{(n)}(0)$, entonces:

$$p_{n+1} = \Phi^{(n+1)}(0) = \Phi(\Phi^{(n)}(0)) = \Phi(p_n)$$

La sucesión (p_n) es monótona creciente pues $p_1 = P[X_1 = 0] = \Phi(0)$ y $p_2 = \Phi(p_1) > \Phi(0) = p_1$. Además, si $p_n > p_{n-1}$, entonces $p_{n+1} = \Phi(p_n) > \Phi(p_{n-1}) = p_n$.

Sea $p = \lim_{n \rightarrow \infty} p_n$, entonces $p = \Phi(p)$.

Sea $r \geq 0$ tal que $r = \Phi(r)$, entonces $p_n = \Phi^{(n)}(0) \leq \Phi^{(n)}(r) = r$, así que $p \leq r$. Por lo tanto, p es la más pequeña solución, en el intervalo $[0, 1]$, de la ecuación $r = \Phi(r)$.

Como $\Phi(1) = 1$ y $\Phi(0) = P[Z = 0] > 0$, $r = 1$ es solución de $r = \Phi(r)$ y $r = 0$ no lo es. Analicemos ahora la función Φ para determinar en que casos existe alguna solución en el intervalo $(0, 1)$. Se tiene:

$$\Phi(t) = E[t^Z] = \sum_{k=0}^{\infty} t^k P[Z = k]$$

$$\Phi'(t) = E[Zt^{Z-1}] = \sum_{k=1}^{\infty} kt^{k-1} P[Z = k]$$

$$\Phi''(t) = E[Z(Z-1)t^{Z-2}] = \sum_{k=2}^{\infty} k(k-1)t^{k-2} P[Z = k]$$

Φ' es continua y diferenciable en el intervalo $(-1, 1)$. Además, como $P[Z = 0] + P[Z = 1] < 1$, entonces $\Phi''(t) > 0$ para cualquier $t \in (0, 1)$, así que Φ' es estrictamente creciente en el intervalo $[0, 1)$.

Sea $f(t) = t - \Phi(t)$. f es entonces continua en el intervalo $[-1, 1]$ y diferenciable en el intervalo $(-1, 1)$. Por otra parte, $f'(t) = 1 - \Phi'(t)$. Así que f' es continua y estrictamente decreciente en el intervalo $[0, 1)$. Además:

$$f'(0) = 1 - \Phi'(0) = 1 - P[Z = 1] > 0.$$

Por lo tanto se tienen los siguientes dos casos:

a) Si $E[Z] = \lim_{t \rightarrow 1^-} \Phi'(t) > 1$, entonces $\lim_{t \rightarrow 1^-} f'(t) < 0$, así que, por el teorema del valor intermedio, existe exactamente un punto $t_0 \in (0, 1)$ tal que $f'(t_0) = 0$. Como $f''(t) = -\Phi''(t) < 0$ para cualquier $t \in (0, 1)$, f , restringida al intervalo $[0, 1]$, alcanza su valor máximo en t_0 y como $f(1) = 0$, entonces $f(t_0) > 0$. Además, $f(0) = -\Phi(0) = -P[Z = 0] < 0$, así que, por el teorema del valor intermedio, existe $r \in (0, 1)$ tal que $f(r) = 0$, es decir $\Phi(r) = r$. Por lo tanto, en este caso se tiene:

$$p = \lim_{n \rightarrow \infty} P[X_n = 0] \in (0, 1)$$

A continuación se ilustran las gráficas de f y de Φ en este caso.

f Φ

b) Si $E[Z] = \lim_{t \rightarrow 1^-} \Phi'(t) \leq 1$, entonces $\lim_{t \rightarrow 1^-} f'(t) \geq 0$, así que $f'(t) > 0$ para cualquier $t \in [0, 1)$, de manera que f es estrictamente creciente en el intervalo $[0, 1]$. Por lo tanto $f(t) < 0$ para cualquier $t \in [0, 1)$, ya que $f(1) = 0$. De manera que no existe $r \in (0, 1)$ tal que $f(r) = 0$. Por lo tanto, en este caso, se tiene:

$$p = \lim_{n \rightarrow \infty} P[X_n = 0] = 1$$

A continuación se ilustran las gráficas de f y de Φ en este caso.

 f Φ

EJEMPLO 4.53 (Función generadora de la suma de un número aleatorio de sumandos). Sea Z_1, Z_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas con función generadora de momentos común Φ y X, Y una pareja de variables aleatorias tales que X toma únicamente valores enteros no negativos, es independiente de Z_1, Z_2, \dots , y, para cada valor k de X , $Y = \sum_{j=1}^k Z_j$. Encuentre la función generadora de Y .

Solución

$$\begin{aligned} E[t^Y | X = k] &= E\left[t^{\sum_{j=1}^k Z_j} | X = k\right] = E\left[t^{\sum_{j=1}^k Z_j}\right] \\ &= \Phi_{Z_1}(t) \cdots \Phi_{Z_k}(t) = [\Phi(t)]^k \end{aligned}$$

Así que:

$$\Phi_Y(t) = E[t^Y] = E(E[t^Y | X]) = E([\Phi(t)]^X) = \Phi_X(\Phi(t))$$

4.7. Distribuciones condicionales en el caso mixto

Si X es una variable aleatoria discreta y Y una variable aleatoria absolutamente continua, no tenemos definida una función de densidad conjunta, así que, para definir una función de densidad condicional, no podemos seguir el mismo procedimiento que seguimos en el caso en que las dos variables aleatorias son, ya sea discretas o absolutamente continuas. Sin embargo, el concepto general de esperanza condicional nos permite definir tanto distribuciones condicionales como distribuciones conjuntas en el caso en que una variable aleatoria es discreta y la otra absolutamente continua.

Sea X una variable aleatoria discreta, con función de densidad f_X , y Y una variable aleatoria absolutamente continua, con función de densidad f_Y . Para cada $x \in \mathbb{R}$ tal que $f_X(x) > 0$ sea $g_x : \mathbb{R} \mapsto \overline{\mathbb{R}}$ una función boreliana no negativa tal que $g_x(Y)$ es una versión de $P[X = x | Y]$.

Si $f_X(x) = 0$, entonces una versión de $P[X = x | Y]$ es la variable aleatoria idénticamente cero, así que, en este caso, se puede tomar g_x idénticamente cero, lo cual asumiremos en lo que sigue.

La idea al definir g_x en la forma mencionada arriba es que, teniendo:

$$g_x(y) = P[X = x | Y = y]$$

h_x es una densidad condicional de X dado que Y toma el valor y .

Sin embargo, como g_x no es única, fijando y , la función $x \mapsto g_x(y)$ no necesariamente es una función de densidad. Esto hace que tengamos que realizar algunos ajustes, los cuales estarán basados en el siguiente resultado:

PROPOSICIÓN 4.54. *La medida del conjunto:*

$$\left\{ y \in \mathbb{R} : f_Y(y) > 0 \text{ y } \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(y) \neq 1 \right\}$$

es cero.

Demostración

Sea B un conjunto boreliano de \mathbb{R} y $x \in \mathbb{R}$ tal que $f_X(x) > 0$, entonces, como $g_x(Y)$ es una versión de $P[X = x | Y]$, se tiene:

$$E[I_B(Y)g_x(Y)] = E[I_B(Y)I_{[X=x]}]$$

Así que, utilizando el teorema de la convergencia monótona:

$$\begin{aligned} E\left[I_B(Y) \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(Y)\right] &= \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} E[I_B(Y)g_x(Y)] \\ &= \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} E[I_B(Y)I_{[X=x]}] = E\left[I_B(Y) \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} I_{[X=x]}\right] \\ &= E[I_B(Y)] \end{aligned}$$

Es decir:

$$\int_B \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(y) f_Y(y) dy = \int_B f_Y(y) dy$$

Por lo tanto:

$$m\left(\left\{y \in \mathbb{R} : f_Y(y) \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(y) \neq f_Y(y)\right\}\right) = 0$$

Finalmente:

$$\begin{aligned} &\left\{y \in \mathbb{R} : f_Y(y) > 0 \text{ y } \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(y) \neq 1\right\} \\ &\subset \left\{y \in \mathbb{R} : f_Y(y) \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} g_x(y) \neq f_Y(y)\right\} \end{aligned}$$

de lo cual se obtiene el resultado. ■

En lo que sigue, C y D serán los conjuntos definidos como sigue:

$$C = \{y \in \mathbb{R} : f_Y(y) > 0\}$$

$$D = \left\{y \in \mathbb{R} : \sum_{\{x \in \mathbb{R} : f_X(x) > 0\}} h_x(y) = 1\right\}$$

Lo que demostramos en la última proposición es que la medida del conjunto $C \cap D^c$ es cero.

V_X denotará al conjunto de posibles valores de X , es decir:

$$V_X = \{x \in \mathbb{R} : f_X(x) > 0\}$$

Para cualquier pareja $(x, y) \in \mathbb{R}^2$, definamos:

$$h_x(y) = \begin{cases} g_x(y) & \text{si } y \in D \\ f_X(x) & \text{en otro caso} \end{cases}$$

$$f_{X|Y}(x | y) = \begin{cases} h_x(y) & \text{si } f_Y(y) > 0 \\ f_X(x) & \text{en otro caso} \end{cases}$$

$$f_{X,Y}(x, y) = f_{X|Y}(x | y)f_Y(y)$$

$$f_{Y|X}(y | x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{si } f_X(x) > 0 \\ f_Y(y) & \text{en otro caso} \end{cases}$$

Obviamente, para cualquier $y \in R$, la función $x \mapsto h_x(y)$ es una función de densidad discreta cuyo conjunto de posibles valores está contenido en el conjunto de posibles valores de X . Además, se tiene el siguiente resultado:

PROPOSICIÓN 4.55. *Para cualquier $x \in \mathbb{R}$, la variable aleatoria $h_x(Y)$ es una versión de $P[X = x | Y]$.*

Demostración

Sea B un conjunto boreliano de \mathbb{R} , entonces:

$$\begin{aligned} E[I_B(Y)h_x(Y)] &= \int_{-\infty}^{\infty} I_B(y)h_x(y)f_Y(y)dy \\ &= \int_{C^c \cup D} I_B(y)h_x(y)f_Y(y)dy = \int_{C^c \cup D} I_B(y)g_x(y)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} I_B(y)g_x(y)f_Y(y)dy = E[I_B(Y)g_x(Y)] = E[I_B(Y)I_{[X=x]}] \end{aligned}$$

Así que $h_x(Y)$ es una versión de $P[X = x | Y]$.

■

PROPOSICIÓN 4.56. *Para cualquier $y \in R$, la función $x \mapsto f_{X|Y}(x | y)$ es una función de densidad discreta cuyo conjunto de posibles valores está contenido en el conjunto de posibles valores de X .*

Demostración

Si $y \in C \cap D$, entonces $f_{X|Y}(x | y) = h_x(y)$ para cualquier $x \in \mathbb{R}$.

Si $y \notin C \cap D$, entonces $f_{X|Y}(x | y) = f_X(x)$ para cualquier $x \in \mathbb{R}$.

Así que, en cualquier caso, $x \mapsto f_{X|Y}(x | y)$ es una función de densidad discreta cuyo conjunto de posibles valores está contenido en el conjunto de posibles valores de X .

■

PROPOSICIÓN 4.57. *Para cualquier $x \in \mathbb{R}$, la función $y \mapsto f_{Y|X}(y | x)$ es una función de densidad.*

Demostración

Si $f_X(x) > 0$, entonces:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{Y|X}(y | x) dy &= \int_{-\infty}^{\infty} \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_C f_{X|Y}(x | y) f_Y(y) dy \\ &= \frac{1}{f_X(x)} \int_C h_x(y) f_Y(y) dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} h_x(y) f_Y(y) dy = \frac{1}{f_X(x)} E[h_x(Y)] \\ &= \frac{1}{f_X(x)} P[X = x] = 1 \end{aligned}$$

Si $f_X(x) = 0$, entonces $y \mapsto f_{Y|X}(y | x)$ es la función f_Y .

Así que, en cualquier caso, $y \mapsto f_{Y|X}(y | x)$ es una función de densidad.

■

PROPOSICIÓN 4.58. *Para cualquier $x \in \mathbb{R}$, se tiene:*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x | y) f_Y(y) dy$$

Demostración

La igualdad $\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} f_{X|Y}(x | y) f_Y(y) dy$ es inmediata.

Por otra parte:

$$\begin{aligned} \int_{-\infty}^{\infty} f_{X|Y}(x | y) f_Y(y) dy &= \int_C f_{X|Y}(x | y) f_Y(y) dy \\ &= \int_C h_x(y) f_Y(y) dy = \int_{-\infty}^{\infty} h_x(y) f_Y(y) dy = E[h_x(Y)] \\ &= P[X = x] = f_X(x) \end{aligned}$$

■

PROPOSICIÓN 4.59. *Para cualquier $y \in \mathbb{R}$, se tiene:*

$$f_Y(y) = \sum_{\{x \in V_X\}} f_{X,Y}(x, y) = \sum_{\{x \in V_X\}} f_{Y|X}(y | x) f_X(x)$$

Demostración

La igualdad $\sum_{\{x \in V_X\}} f_{X,Y}(x, y) = \sum_{\{x \in V_X\}} f_{Y|X}(y | x) f_X(x)$ es inmediata.

Por otra parte, si $f_Y(y) > 0$, entonces:

$$\begin{aligned} \sum_{\{x \in V_X\}} f_{X,Y}(x, y) &= \sum_{\{x \in V_X\}} f_{X|Y}(x | y) f_Y(y) \\ &= f_Y(y) \sum_{\{x \in V_X\}} h_x(y) = f_Y(y) \end{aligned}$$

Si $f_Y(y) = 0$, entonces:

$$\sum_{\{x \in V_X\}} f_{X,Y}(x, y) = \sum_{\{x \in V_X\}} f_{X|Y}(x | y) f_Y(y) = 0$$

■

COROLARIO 4.60. $\sum_{\{x \in V_X\}} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} \sum_{\{x \in V_X\}} f_{X,Y}(x, y) dy = 1$

PROPOSICIÓN 4.61. Para cualesquiera $x, a, b \in \mathbb{R}$, con $a < b$, se tiene:

$$P[X = x, a < Y < b] = \int_a^b f_{X,Y}(x, y) dy$$

Demostración

$$\begin{aligned} \int_a^b f_{X,Y}(x, y) dy &= \int_a^b f_{X|Y}(x | y) f_Y(y) dy = \int_{-\infty}^{\infty} I_{(a,b)}(y) f_{X|Y}(x | y) f_Y(y) dy \\ &= \int_C I_{(a,b)}(y) f_{X|Y}(x | y) f_Y(y) dy = \int_C I_{(a,b)}(y) h_x(y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} I_{(a,b)}(y) h_x(y) f_Y(y) dy = E[I_{(a,b)}(Y) h_x(Y)] \\ &= E[I_{(a,b)}(Y) I_{[X=x]}] = E[I_{[a < Y < b]} I_{[X=x]}] = E[I_{[a < Y < b, X=x]}] \\ &= P[X = x, a < Y < b] \end{aligned}$$

■

COROLARIO 4.62. Para cualquier pareja de conjuntos borelianos A y B de \mathbb{R} , se tiene:

$$P[X \in A, Y \in B] = \sum_{\{x \in A \cap V_X\}} \int_B f_{X,Y}(x, y) dy = \int_B \sum_{\{x \in A \cap V_X\}} f_{X,Y}(x, y) dy$$

COROLARIO 4.63. X y Y son independientes si y sólo si el conjunto:

$$\{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x) f_Y(y) \text{ para alguna } x \in \mathbb{R}\}$$

tiene medida cero.

Demostración

Primero obsérvese que si $f_X(x) = 0$ ó $f_Y(y) = 0$, entonces $f_{X,Y}(x, y) = 0$, así que $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Supongamos que X y Y son independientes, entonces,

para cualesquiera $x, a, b \in \mathbb{R}$, con $a < b$, se tiene:

$$\begin{aligned} \int_a^b f_{X,Y}(x, y)dy &= P[X = x, a < Y < b] = P[X = x]P[a < Y < b] \\ &= f_X(x) \int_a^b f_Y(y)dy = \int_a^b f_X(x)f_Y(y)dy \end{aligned}$$

Así que la medida del conjunto:

$$\{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x)f_Y(y)\}$$

es cero.

Por lo tanto, el conjunto:

$$\begin{aligned} \{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x)f_Y(y) \text{ para alguna } x \in \mathbb{R}\} \\ = \bigcup_{\{x \in V_X\}} \{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x)f_Y(y)\} \end{aligned}$$

también tiene medida cero.

Supongamos ahora que el conjunto:

$$\begin{aligned} E = \{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x)f_Y(y) \text{ para alguna } x \in \mathbb{R}\} \\ = \bigcup_{\{x \in V_X\}} \{y \in \mathbb{R} : f_{X,Y}(x, y) \neq f_X(x)f_Y(y)\} \end{aligned}$$

tiene medida cero.

Entonces, para cualquier pareja de conjuntos borelianos A y B de \mathbb{R} , se tiene:

$$\begin{aligned} P[X \in A, Y \in B] &= \sum_{\{x \in A \cap V_X\}} \int_B f_{X,Y}(x, y)dy \\ &= \sum_{\{x \in A \cap V_X\}} \int_{B \cap E^c} f_{X,Y}(x, y)dy \\ &= \sum_{\{x \in A \cap V_X\}} \int_{B \cap E^c} f_X(x)f_Y(y)dy = \sum_{\{x \in A \cap V_X\}} f_X(x) \int_B f_Y(y)dy \\ &= P[X \in A]P[Y \in B] \end{aligned}$$

Así que X y Y son independientes. ■

PROPOSICIÓN 4.64. Sea $g : \mathbb{R} \mapsto \mathbb{R}$ una función boreliana tal que $g(X)$ tiene esperanza finita. Definamos la función $h : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:

$$h(y) = \begin{cases} \sum_{\{x \in V_X\}} g(x) f_{X|Y}(x | y) & \text{si } \sum_{\{x \in V_X\}} |g(x)| f_{X|Y}(x | y) < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Entonces $h(Y)$ es una versión de la esperanza condicional $E[g(X) | Y]$.

Demostración

$$g(x) = I_A(x)$$

$$\begin{aligned} E[I_B(Y)h(Y)] &= \int_{-\infty}^{\infty} I_B(y)h(y)f_Y(y)dy \\ &= \int_{-\infty}^{\infty} I_B(y)f_Y(y) \sum_{\{x \in V_X\}} I_A(x)f_{X|Y}(x | y)dy \\ &= \sum_{\{x \in V_X\}} I_A(x) \int_C I_B(y)h_x(y)f_Y(y)dy \\ &= \sum_{\{x \in V_X\}} I_A(x) \int_{-\infty}^{\infty} I_B(y)h_x(y)f_Y(y)dy \\ &= \sum_{\{x \in V_X\}} I_A(x)E[I_B(Y)h_x(Y)] = \sum_{\{x \in V_X\}} I_A(x)E[I_B(Y)I_{[X=x]}] \\ &= \sum_{\{x \in A \cap V_X\}} E[I_B(Y)I_{[X=x]}] = E\left[I_B(Y) \sum_{\{x \in A \cap V_X\}} I_{[X=x]}\right] \\ &= E[I_B(Y)I_{A \cap V_X}(X)] = E[I_A(X)I_B(Y)] - E[I_{A \cap V_X^c}(X)I_B(Y)] \\ &= E[I_B(Y)I_A(X)] = E[I_B(Y)g(X)] \end{aligned}$$

ya que:

$$E[I_{A \cap V_X^c}(X)I_B(Y)] \leq E[I_{V_X^c}(X)] = P[X \in V_X^c] = 0$$

■

PROPOSICIÓN 4.65. Sea $g : \mathbb{R} \mapsto \mathbb{R}$ una función boreliana tal que $g(Y)$ tiene esperanza finita. Definamos la función $h : \mathbb{R} \mapsto \mathbb{R}$ de la siguiente manera:

$$h(x) = \begin{cases} \int_{-\infty}^{\infty} g(y)f_{Y|X}(y | x)dy & \text{si } \int_{-\infty}^{\infty} |g(y)| f_{Y|X}(y | x)dy < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Entonces $h(X)$ es una versión de la esperanza condicional $E[g(Y) | X]$.

Demostración

$$g(y) = I_A(y)$$

$$\begin{aligned}
E[I_B(X)h(X)] &= \sum_{\{x \in V_X\}} I_B(x)h(x)f_X(x) \\
&= \sum_{\{x \in V_X\}} I_B(x)f_X(x) \int_{-\infty}^{\infty} I_A(y)f_{Y|X}(y|x)dy \\
&= \sum_{\{x \in V_X\}} I_B(x) \int_{-\infty}^{\infty} I_A(y)f_{Y|X}(y|x)f_X(x)dy \\
&= \sum_{\{x \in V_X\}} I_B(x) \int_{-\infty}^{\infty} I_A(y)f_{X|Y}(x|y)f_Y(y)dy \\
&= \sum_{\{x \in V_X\}} I_B(x) \int_{-\infty}^{\infty} I_A(y)h_x(y)f_Y(y)dy \\
&= \sum_{\{x \in V_X\}} I_B(x)E[I_A(Y)h_x(Y)] = \sum_{\{x \in V_X\}} I_B(x)E[I_A(Y)I_{[X=x]}] \\
&= \sum_{\{x \in B \cap V_X\}} E[I_A(Y)I_{[X=x]}] = E\left[I_A(Y) \sum_{\{x \in B \cap V_X\}} I_{[X=x]}\right] \\
&= E[I_A(Y)I_{B \cap V_X}(X)] = E[I_B(X)I_A(Y)] - E[I_{B \cap V_X^c}(X)I_A(Y)] \\
&= E[I_B(X)I_A(Y)] = E[I_B(X)g(Y)]
\end{aligned}$$

ya que:

$$E[I_{B \cap V_X^c}(X)I_A(Y)] \leq E[I_{V_X^c}(X)] = P[X \in V_X^c] = 0$$

■

DEFINICIÓN 4.66. Para cualquier $y \in \mathbb{R}$, la función $x \mapsto f_{X|Y}(x|y)$ será llamada la función de densidad condicional de X , dado que $Y = y$, y a la distribución que define, la distribución condicional de X dado que $Y = y$.

DEFINICIÓN 4.67. Para cualquier $x \in \mathbb{R}$, la función $y \mapsto f_{Y|X}(y|x)$ será llamada la función de densidad condicional de Y , dado que $X = x$, y a la distribución que define, la distribución condicional de Y dado que $X = x$.

DEFINICIÓN 4.68. La función $f_{X,Y}$ será llamada la función de densidad conjunta de la pareja (X, Y) .

Obsérvese que, si $f_X(x) > 0$, se tiene:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Esta relación puede verse como una fórmula de Bayes. Mediante ella se puede obtener la distribución condicional de Y , dado que $X = x$, a partir de la distribución condicional de X dado que $Y = y$. Esto no es casual pues, en un caso particular, esta fórmula está implícita en un resultado de Thomas Bayes, el cual motiva que el método que consiste en calcular una probabilidad condicional $P(A|B)$ a partir de $P(B|A)$ se conozca como la regla de Bayes. Éste en realidad no demostró la regla para el caso

de dos eventos A y B , la cual estaba ya implícita, por lo menos, en el trabajo previo de Abraham de Moivre. La aportación de Bayes es en realidad más significativa pues se refiere a un problema de distribuciones mixtas. El resultado original de Bayes se trata en el siguiente ejemplo:

EJEMPLO 4.69. *En el año 1763 se publicó un artículo de Thomas Bayes¹ en el cual se plantea y resuelve el siguiente problema:*

Dado el número de veces en el cual un evento desconocido ha ocurrido y fallado, encontrar la probabilidad (chance) de que su probabilidad de ocurrencia en un ensayo esté comprendida entre dos valores dados.

Para resolver este problema, Bayes consideró un plano $ABCD$, el cual está hecho de tal manera que si una bola es lanzada sobre él, entonces: habrá la misma probabilidad de que permanezca en cualquiera de dos partes iguales del plano y necesariamente permanecerá sobre éste.

Una bola \mathcal{W} es lanzada primero y, a través del punto en donde cae, se traza una recta paralela a AD , la cual corta al segmento AB en el punto s .

C

D

A

s

B

Después de lanzar la bola \mathcal{W} , se lanza una bola \mathcal{O} n veces sobre el plano. En cada lanzamiento se dirá que el evento M ocurre si la bola \mathcal{O} cae en el rectángulo sA .

Bayes mostró que cuando n está dado, antes de que la bola \mathcal{W} sea lanzada, lo cual determina la probabilidad de ocurrencia del evento M , la probabilidad de que el evento M ocurra k veces en los n ensayos es la misma para cualquier k . Argumentaba también que el evento cuya probabilidad de ocurrencia se quiere estimar tiene la misma propiedad pues, antes de disponer de información sobre el número de veces que ocurre en n ensayos, no hay razón para pensar que, en un cierto número de ensayos, debería

¹An essay towards solving a problem in the doctrine of chances, Philos. Trans. Roy. Soc. London, Ser. A, 53, 1763. Reproducido en Biometrika 45, 1958.

ocurrir algún número de veces en lugar de otro. Con base en esto concluyó que el problema planteado originalmente se puede resolver encontrando la probabilidad de ocurrencia del evento M sabiendo que éste ocurre k veces en los n ensayos. Si llamamos p a la probabilidad de ocurrencia del evento M , X al número de ocurrencias del evento M en los n ensayos y $0 < a < b < 1$, entonces Bayes llegó a los siguientes resultados:

$$P(X = k \mid p = y) = \binom{n}{k} y^k (1 - y)^{n-k}$$

$$P(X = k, a < p < b) = \int_a^b P(X = k \mid p = y) dy = \int_a^b \binom{n}{k} y^k (1 - y)^{n-k} dy$$

$$P(X = k) = \int_0^1 P(X = k \mid p = y) dy = \int_0^1 \binom{n}{k} y^k (1 - y)^{n-k} dy$$

$$P[a < p < b \mid X = k] = \frac{P(X=k, a < p < b)}{P(X=k)} = \frac{\int_a^b \binom{n}{k} y^k (1-y)^{n-k} dy}{\int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy}$$

En lenguaje moderno, lo que hizo Bayes fue resolver un problema de distribuciones condicionales. Si p es la probabilidad de ocurrencia de un evento al realizar un cierto experimento aleatorio y X es el número de veces en los cuales el evento ocurre al repetir n veces el experimento, Bayes se planteó el problema de encontrar la distribución condicional de p dado que $X = k$, asumiendo que p es una cantidad que originalmente se selecciona al azar en el intervalo $(0, 1)$.

Conociendo el valor de p , X tiene distribución binomial con parámetros n y p . De manera que el problema de Bayes equivale al siguiente:

Sea Y una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución binomial de parámetros n y $p = y$. Encontrar la distribución condicional de Y dado que $X = k$, para $k \in \{0, \dots, n\}$.

Este problema, de acuerdo con los resultados demostrados arriba, se resuelve de la siguiente manera:

Dados $y \in (0, 1)$ y $k \in \{0, \dots, n\}$, se tiene:

$$f_{X|Y}(k \mid y) = P[X = k \mid Y = y] = \binom{n}{k} y^k (1 - y)^{n-k}$$

Por lo tanto:

$$P[X = k] = \int_{-\infty}^{\infty} f_{X|Y}(k \mid y) f_Y(y) dy = \int_0^1 \binom{n}{k} y^k (1 - y)^{n-k} dy$$

$$f_{Y|X}(y \mid k) = \frac{f_{X|Y}(k|y) f_Y(y)}{P[X=k]} = \frac{\binom{n}{k} y^k (1-y)^{n-k}}{\int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy}$$

Es decir, dado que $X = k$, Y tiene distribución beta con parámetros $k + 1$ y $n - k + 1$.

Así que, si $0 < a < b < 1$, se tiene:

$$P[a < Y < b \mid X = k] = \int_a^b f_{Y|X}(y \mid k) dy = \frac{\int_a^b \binom{n}{k} y^k (1-y)^{n-k} dy}{\int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy}$$

De esta manera tenemos lo que se puede llamar el **teorema de Bayes**:

Cuando lo único que se conoce de un evento A es que ha ocurrido k veces y fallado $n - k$ en n ensayos, entonces su probabilidad de ocurrencia puede considerarse como seleccionada de una población Y con distribución beta de parámetros $k + 1$ y $n - k + 1$.

Obsérvese que, efectivamente, como lo afirmó Bayes, antes de conocer el valor de p , la probabilidad de que el evento M ocurra k veces en los n ensayos, es decir $P[X = k]$, es la misma para cualquier k . Específicamente, se tiene:

$$\begin{aligned} P[X = k] &= \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} dy = \binom{n}{k} B(k+1, n-k+1) \\ &= \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1} \end{aligned}$$

Este último resultado equivale a decir que si Y es una variable aleatoria con distribución uniforme en el intervalo $[0, 1]$ y, dado que $Y = y$, X tiene distribución binomial con parámetros n y $p = y$, entonces X tiene distribución uniforme en el conjunto $\{0, \dots, n\}$.

EJEMPLO 4.70. Supongamos que el número de accidentes que tiene una persona en un año tiene distribución Poisson con parámetro Y , de tal manera que, para cada $y > 0$, el porcentaje de personas para las cuales $Y > y$ es igual a $e^{-\lambda y}$, en donde λ es una constante positiva. Si X es el número de accidentes en un año de una persona seleccionada al azar, encuentre a) la distribución de X , b) $E[X]$, c) la distribución condicional de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$, y d) $E[Y \mid X]$.

Solución

a. Para $x \in \{0, 1, \dots\}$, se tiene:

$$\begin{aligned} P[X = x] &= \int_0^\infty P[X = x \mid Y = y] f_Y(y) dy = \int_0^\infty \frac{e^{-y} y^x}{x!} \lambda e^{-\lambda y} dy \\ &= \frac{\lambda}{x!} \int_0^\infty y^x e^{-(\lambda+1)y} dy = \frac{\lambda}{(\lambda+1)^{x+1}} \int_0^\infty \frac{(\lambda+1)^{x+1} y^x e^{-(\lambda+1)y}}{x!} dy = \frac{\lambda}{(\lambda+1)^{x+1}} \\ &= \left(\frac{\lambda}{\lambda+1}\right) \left(\frac{1}{\lambda+1}\right)^x \end{aligned}$$

Por lo tanto, X tiene distribución geométrica con parámetro $p = \frac{\lambda}{\lambda+1}$.

$$b. E[X] = \int_{-\infty}^{\infty} E[X | Y = y] f_Y(y) dy = \int_0^{\infty} y \lambda e^{-\lambda y} dy = \frac{1}{\lambda}$$

c. Para $x \in \{0, 1, \dots\}$ y $y > 0$, se tiene:

$$f_{Y|X}(y | x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{\frac{e^{-y}y^x}{x!} \lambda e^{-\lambda y}}{\left(\frac{\lambda}{\lambda+1}\right)\left(\frac{1}{\lambda+1}\right)^x} = \frac{(\lambda+1)^{x+1} y^x e^{-(\lambda+1)y}}{x!}$$

d. De acuerdo con el resultado en la parte c, si $x \in \{0, 1, \dots\}$, entonces, dado que $X = x$, Y tiene distribución gama con parámetros $x + 1$ y $\lambda + 1$. Así que, $E[Y | X] = \frac{x+1}{\lambda+1}$.

EJEMPLO 4.71. Supongamos que se tienen N especies animales, de tal manera que el número de individuos de la especie i que son atrapados en una determinada trampa tiene distribución Poisson con parámetro Y_i . Supongamos además que Y_1, \dots, Y_N son variables aleatorias independientes, todas ellas con distribución gama de parámetros α y λ . Encuentre el número esperado de especies que se encuentren representadas en la trampa con al menos un individuo.

Solución

Sea X_i el número de individuos de la especie i que se encuentren en la trampa, Z_k el número de especies que se encuentren representadas en la trampa con k individuos y Z el número de especies que se encuentren representadas en la trampa con al menos un individuo. Definamos además, para $i \in \{1, \dots, N\}$ y $k \in \{0, 1, \dots\}$:

$$Z_k^i = \begin{cases} 1 & \text{si } X_i = k \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene entonces:

$$E[Z_k^i | Y_i = y] = P[X_i = k | Y_i = y] = \frac{1}{k!} y^k e^{-y}$$

Así que:

$$\begin{aligned} E[Z_k^i] &= \frac{1}{k!} \int_0^{\infty} y^k e^{-y} \frac{\lambda^\alpha y^{\alpha-1} e^{-\lambda y}}{\Gamma(\alpha)} dy = \frac{\lambda^\alpha}{k! \Gamma(\alpha)} \int_0^{\infty} y^{k+\alpha-1} e^{-y(1+\lambda)} dy \\ &= \frac{\lambda^\alpha}{k! \Gamma(\alpha)} \frac{1}{(1+\lambda)^{k+\alpha}} \int_0^{\infty} t^{k+\alpha-1} e^{-t} dt = \frac{\lambda^\alpha}{k! \Gamma(\alpha)} \frac{\Gamma(k+\alpha)}{(1+\lambda)^{k+\alpha}} = \frac{\Gamma(k+\alpha)}{k! \Gamma(\alpha)} \left(\frac{\lambda}{1+\lambda}\right)^\alpha \left(\frac{1}{1+\lambda}\right)^k \end{aligned}$$

Además, $Z_k = Z_k^1 + Z_k^2 + \dots + Z_k^N$, así que:

$$E[Z_k] = N \frac{\Gamma(k+\alpha)}{k! \Gamma(\alpha)} \left(\frac{\lambda}{1+\lambda}\right)^\alpha \left(\frac{1}{1+\lambda}\right)^k$$

Finalmente, $Z = \sum_{k=1}^{\infty} Z_k$, así que:

$$E[Z] = \sum_{k=1}^{\infty} E[Z_k] = N \sum_{k=1}^{\infty} \frac{\Gamma(k+\alpha)}{k! \Gamma(\alpha)} \left(\frac{\lambda}{1+\lambda}\right)^\alpha \left(\frac{1}{1+\lambda}\right)^k$$

Pero los términos $\frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \left(\frac{\lambda}{1+\lambda}\right)^\alpha \left(\frac{1}{1+\lambda}\right)^k$ corresponden a los de una distribución binomial negativa con parámetros α y $p = \frac{\lambda}{1+\lambda}$, Por lo tanto:

$$E[Z] = N \left[1 - \left(\frac{\lambda}{\lambda+1}\right)^r\right]$$

EJEMPLO 4.72. Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t , entonces la familia de variables aleatorias $\{X_t\}_{t \geq 0}$ forma un proceso de Poisson de parámetro λ . Vamos a encontrar la distribución conjunta de T_1, \dots, T_n , dado que $X_t = n$, en donde $t > 0$ y $n \in \mathbb{R}$.

Recordemos que:

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \begin{cases} \lambda^n e^{-\lambda t_n} & \text{si } 0 < t_1 < \dots < t_n \\ 0 & \text{en otro caso} \end{cases}$$

$$= \lambda^n e^{-\lambda t_n} I_{\{0 < y_1 < \dots < y_n\}}(t_1, \dots, t_n)$$

Sean $0 < t_1 < \dots < t_n$, entonces:

$$\begin{aligned} f_{T_1, \dots, T_n | X_t}(t_1, \dots, t_n | n) &= \frac{1}{P[X_t=n]} P[T_1 \leq t_1, \dots, T_n \leq t_n, X_t = n] \\ &= \frac{1}{P[X_t=n]} P[T_1 \leq t_1, \dots, T_n \leq t_n, T_n \leq t, T_{n+1} > t] \\ &= \frac{1}{P[X_t=n]} \int \dots \int_{\{x_1 \leq t_1, \dots, x_n \leq t_n\}} \int_t^\infty \lambda^{n+1} e^{-\lambda x_{n+1}} I_{\{0 < y_1 < \dots < y_{n+1}\}}(x_1, \dots, x_n, t) dx_{n+1} \dots dx_1 \\ &= \frac{n!}{\lambda^n t^n e^{-\lambda t}} \int \dots \int_{\{x_1 \leq t_1, \dots, x_n \leq t_n\}} I_{\{0 < y_1 < \dots < y_{n+1}\}}(x_1, \dots, x_n, t) \int_t^\infty \lambda^{n+1} e^{-\lambda x_{n+1}} dx_{n+1} \dots dx_1 \\ &= \frac{n!}{\lambda^n t^n e^{-\lambda t}} \lambda^n e^{-\lambda t} \int \dots \int_{\{x_1 \leq t_1, \dots, x_n \leq t_n\}} I_{\{0 < y_1 < \dots < y_{n+1}\}}(x_1, \dots, x_n, t) dx_n \dots dx_1 \\ &= \int \dots \int_{\{x_1 \leq t_1, \dots, x_n \leq t_n\}} I_{\{0 < y_1 < \dots < y_{n+1}\}}(x_1, \dots, x_n, t) \frac{n!}{t^n} dx_n \dots dx_1 \end{aligned}$$

Así que:

$$\begin{aligned} f_{T_1, \dots, T_n | X_t}(t_1, \dots, t_n | n) &= I_{\{0 < y_1 < \dots < y_{n+1}\}}(t_1, \dots, t_n, t) \frac{n!}{t^n} \\ &= \begin{cases} \frac{n!}{t^n} & \text{si } 0 < t_1 < \dots < t_n < t \\ 0 & \text{en otro caso} \end{cases} \end{aligned}$$

Por lo tanto, dado que $X_t = n$, la distribución conjunta de T_1, \dots, T_n es la misma que la de los estadísticos de orden correspondientes a n variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, t)$.

EJERCICIOS

EJERCICIO 4.1. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, 2, \dots, 2N\}$, en donde N es un entero mayor que 1. Encuentre a) $E[X | X > 2Y]$ y b) $E[X | X + Y > 4]$.

EJERCICIO 4.2. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre:

a) $E[X | X > 2Y]$

b) $E[X | X < 2Y]$

EJERCICIO 4.3. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro $\lambda = 2$. Encuentre a) $E[X | X > 2Y + 1]$ y b) $E[X | X > 2Y - 1]$.

EJERCICIO 4.4. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $E[X | Y > 2X]$ y b) $E[X | X + Y > 1]$.

EJERCICIO 4.5. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(-1, 1)$. Encuentre a) $E[X | X < Y^2]$ y b) $E[Y | X < Y^2]$.

EJERCICIO 4.6. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{N^2(N+1)}(x + y) & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[X | Y]$.

EJERCICIO 4.7. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{N(N^2-1)}(y - x) & \text{si } x < y \text{ y } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $E[X | Y]$ y b) $E[Y | X]$.

EJERCICIO 4.8. Se eligen, al azar y sin reemplazo, dos tarjetas de una urna que contiene N tarjetas numeradas del 1 al N , en donde N es un entero mayor que 1. Sean X y Y el menor y mayor, respectivamente, de los números de las tarjetas seleccionadas. Encuentre a) $E[X | Y]$ y b) $E[Y | X]$.

EJERCICIO 4.9. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \dots, N\}$. Encuentre a) $E[X | Y - X]$ y b) $E[Y | Y - X]$.

EJERCICIO 4.10. Sean X y Y dos variables aleatorias independientes, ambas con distribución geométrica de parámetro p . Encuentre $E[X | \min(X, Y)]$.

EJERCICIO 4.11. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \dots, N\}$. Encuentre $E[X | \max(X, Y)]$.

EJERCICIO 4.12. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \dots, N\}$. Encuentre a) $E[\min(X, Y) | Y]$ y b) $E[\max(X, Y) | Y]$.

EJERCICIO 4.13. Demuestre i, ii y iii de la proposición 4.14.

EJERCICIO 4.14. Sea X una variable aleatoria de esperanza finita y Y cualquier variable aleatoria. Demuestre que $|E[X | Y]| \leq E[|X| | Y]$.

EJERCICIO 4.15. Sean X_1, \dots, X_n n variables aleatorias independientes, idénticamente distribuidas y de esperanza finita. Demuestre que:

$$E\left[X_k \mid \sum_{j=1}^n X_j\right] = \frac{1}{n} \sum_{j=1}^n X_j$$

para cualquier $k \in \{1, \dots, n\}$.

EJERCICIO 4.16. Sean X y Y dos variables aleatorias de esperanza y varianza finitas. Demuestre que $\text{Cov}(X, E[Y | X]) = \text{Cov}(X, Y)$.

EJERCICIO 4.17. Sean X y Y dos variables aleatorias de esperanza finita tales que $E[Y | X] = E[Y]$. Asumiendo que XY tiene esperanza finita, demuestre que $\text{Cov}(X, Y) = 0$.

EJERCICIO 4.18. Sean X , Y y Z tres variables aleatorias tales que X tiene esperanza finita y Y está acotada. Demuestre que a) $E[Y | Z]$ está acotada y b) $E[YE(X | Z)] = E[XE(Y | Z)]$.

EJERCICIO 4.19. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Sea $Z = X + Y$ y definamos:

$$h(z) = \begin{cases} \int_{-\infty}^{\infty} x \frac{f_{X,Z}(x,z)}{f_Z(z)} & \text{si } f_Z(z) > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Demuestre que $h(Z)$ es una versión de la esperanza condicional $E[X | Z]$.

EJERCICIO 4.20. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{8}(y^2 - x^2)e^{-y} & \text{si } -y \leq x \leq y, \quad 0 < y < \infty \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $E[X | Y]$ y b) $E[Y | X]$.

EJERCICIO 4.21. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{8}ye^{-y} & \text{si } 0 < x < y \\ \frac{1}{8}y^2e^y & \text{si } y < x < 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre: a) $E[X | Y]$ y b) $E[Y | X]$.

EJERCICIO 4.22. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{11}(4 - x - y) & \text{si } 0 < x < 1, \quad 0 < y < x + 2 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[(X + Y)^2 | Y]$.

EJERCICIO 4.23. Sean X y Y dos variables aleatorias independientes, con funciones de densidad dadas por:

$$f_X(x) = \begin{cases} \frac{1}{2}\alpha e^{-\alpha x} & \text{si } x > 0 \\ \frac{1}{2}\alpha e^{\alpha x} & \text{si } x \leq 0 \\ 0 & \text{en otro caso} \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{2}\beta e^{-\beta y} & \text{si } y > 0 \\ \frac{1}{2}\beta e^{\beta y} & \text{si } y \leq 0 \\ 0 & \text{en otro caso} \end{cases}$$

respectivamente, en donde α y β son constantes positivas. Encuentre $E[(X + Y)^3 | Y]$.

EJERCICIO 4.24. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \frac{1}{y}e^{-\frac{x}{y}}e^{-y} & \text{si } x > 0, \quad y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[e^{-(X+Y)} | Y]$.

EJERCICIO 4.25. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda x} & \text{si } 0 < y < x \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $E[X^2 | Y]$ y b) $E[X | Y^2]$.

EJERCICIO 4.26. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(-1, 1)$. Encuentre a) $E[XY | X]$ y b) $E[X | XY]$.

EJERCICIO 4.27. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) $E[Y | \frac{X}{Y}]$ y b) $E[\frac{X}{Y} | Y]$.

EJERCICIO 4.28. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[X + Y | Y - X]$.

EJERCICIO 4.29. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre $E[\max(X, Y) | X]$.

EJERCICIO 4.30. Sean X y Y dos variables aleatorias independientes, ambas con distribución gama, X de parámetros α y λ , Y de parámetros β y λ . Encuentre $E[X + Y | \frac{X}{X+Y}]$.

EJERCICIO 4.31. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre $E[XY | X + Y]$.

EJERCICIO 4.32. Sea X una variable aleatoria con distribución exponencial de parámetro $\lambda = 1$, $t > 0$, $U = \min\{X, t\}$ y $V = \max\{X, t\}$. Encuentre $E[X | U]$ y $E[X | V]$.

EJERCICIO 4.33. Sean X y Y las coordenadas de un punto que se elige al azar en el interior del círculo de radio 1 y centro en el origen. Encuentre $E[X | Y]$.

EJERCICIO 4.34. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{N^2(N+1)}(x + y) & \text{si } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

Para $y \in \{1, \dots, N\}$, encuentre la función de densidad condicional de X dado que $Y = y$.

EJERCICIO 4.35. Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{N(N^2-1)}(y-x) & \text{si } x < y \text{ y } x, y \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

a) Para $y \in \{2, \dots, N\}$, encuentre la función de densidad condicional de X dado que $Y = y$ y b) para $x \in \{1, \dots, N-1\}$, encuentre la función de densidad condicional de Y dado que $X = x$.

EJERCICIO 4.36. Se eligen, al azar y sin reemplazo, dos tarjetas de una urna que contiene N tarjetas numeradas del 1 al N , en donde N es un entero mayor que 1. Sean X y Y el menor y mayor, respectivamente, de los números de las tarjetas seleccionadas. Encuentre la función de densidad condicional de a) X dado que $Y = y$, para $y \in \{2, \dots, N\}$ y b) Y dado que $X = x$, para $x \in \{1, \dots, N-1\}$.

EJERCICIO 4.37. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \dots, N\}$. Para $x, y \in \{1, \dots, N\}$, encuentre la función de densidad condicional de a) $\max(X, Y)$ dado que $Y = y$ y b) $\min(X, Y)$ dado que $X = x$.

EJERCICIO 4.38. Consideremos una sucesión de ensayos de Bernoulli independientes en cada uno de los cuales la probabilidad de éxito es igual a p y, para $k \in \mathbb{N}$, sea X_k el número de ensayo en el cual ocurre el k -ésimo éxito. Para $n \in \{2, 3, \dots\}$, encuentre la función de densidad condicional de X_1 dado que $X_2 = n$.

EJERCICIO 4.39. Sean X y Y dos variables aleatorias independientes, ambas con distribución Poisson, X de parámetro λ_1 y Y de parámetro λ_2 . Para $z \in \{0, 1, \dots\}$, encuentre e identifique la densidad condicional de X dado que $X + Y = z$.

EJERCICIO 4.40. Sea (N_1, \dots, N_r) un vector aleatorio con distribución multinomial de parámetros n, p_1, \dots, p_r . a) Encuentre la distribución de N_j dado que $N_i = s$, para $i, j \in \{1, \dots, r\}$, $i \neq j$ y $s \in \{0, \dots, n\}$. b) Utilice el resultado de la parte a para calcular $\text{Cov}(N_i, N_j)$.

EJERCICIO 4.41. Un experimento aleatorio consiste en seleccionar al azar un punto en el interior del triángulo de vértices $(0, 0)$, $(2, 0)$ y $(1, 2)$. Sean X y Y la abscisa y ordenada, respectivamente, del punto seleccionado. Encuentre la distribución condicional de Y dada X .

EJERCICIO 4.42. Sean X_1, X_2, X_3 tres variables aleatorias independientes, las 3 con distribución uniforme en el intervalo $(0, 1)$ y sean $X_{(1)}, X_{(2)}, X_{(3)}$ los estadísticos de orden correspondientes a X_1, X_2, X_3 . Encuentre a) una función de densidad condicional de $X_{(3)}$ dado que $X_{(1)} = x_1$, para $x_1 \in (0, 1)$ y b) $E[X_{(3)} | X_{(1)}]$.

EJERCICIO 4.43. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 1)$ y coeficiente de correlación $\frac{1}{2}$. Encuentre a) una función de densidad conjunta de $U = X + 2Y$ y $V = 2X - Y$ y b) $E[U | V]$.

EJERCICIO 4.44. Sean X y Y dos variables aleatorias, ambas con distribución normal estándar, tales que la distribución conjunta de X y Y es normal bivariada con coeficiente de correlación $\rho = \frac{1}{3}$. Encuentre $P[-1 < Y < 1 \mid X = 1]$.

EJERCICIO 4.45. Sean X , Y y Z tres variables aleatorias independientes, todas con distribución normal estándar. Encuentre $E[2X + Y + Z \mid X + Y - Z]$ y $E[(2X + Y + Z)^2 \mid X + Y - Z]$.

EJERCICIO 4.46. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Para $z \in (0, 2)$, encuentre la distribución condicional de X dado que $X + Y = z$.

EJERCICIO 4.47. Sean X y Y dos variables aleatorias independientes, ambas con distribución normal estándar. Para $z \in \mathbb{R}$, encuentre la distribución condicional de X dado que $X + Y = z$.

EJERCICIO 4.48. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f_{X,Y}(x, y) = \begin{cases} 8xy & \text{si } 0 < x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $E[X \mid Y]$, b) una función de densidad condicional de X dado que $Y - X = z$, para cualquier $z \in \mathbb{R}$ tal que $f_{Y-X}(z) > 0$ y c) $E[X \mid Y - X < \frac{1}{4}]$.

EJERCICIO 4.49. Sean X y Y dos variables aleatorias con función de densidad conjunta dada por:

$$f(x, y) = \begin{cases} 6x & \text{si } 0 < x < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Encuentre a) $E[X \mid Y]$, b) $E[X \mid Y - X]$, c) $E[X \mid Y - X < \frac{1}{4}]$ y d) $E[Ye^{X-Y} \mid Y]$.

EJERCICIO 4.50. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . Encuentre a) la distribución condicional de Y dada $Y - X$ y b) $E[Y \mid Y - X]$.

EJERCICIO 4.51. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Para $z \in (-1, 1)$, encuentre las distribuciones condicionales de X y Y dado que $Y - X = z$.

EJERCICIO 4.52. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 1)$ y coeficiente de correlación $\frac{1}{2}$. Para $z \in \mathbb{R}$, encuentre e identifique la distribución condicional de X dado que $Y - X = z$.

EJERCICIO 4.53. Sea (X, Y) un vector aleatorio con distribución normal bivariada con vector de esperanzas $(0, 0)$, vector de varianzas $(1, 4)$ y coeficiente de correlación

$\frac{1}{4}$. a) Encuentre una función de densidad conjunta de la pareja $X, Y - 4X$. b) Para $z \in \mathbb{R}$, encuentre e identifique la distribución condicional de X dado que $Y - 4X = z$.

EJERCICIO 4.54. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$. Encuentre a) la función de densidad condicional de $X + Y$ dado que $Y - X = v$, para $v \in (-1, 1)$ y b) $E[X + Y | Y - X]$.

EJERCICIO 4.55. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el intervalo $(0, 1)$ y sea $Z = \max(X, Y)$. Para $y \in (0, 1)$, encuentre la distribución condicional de Z dado que $Y = y$, y utilícela para calcular $E(Z | Y)$.

EJERCICIO 4.56. Sean X y Y dos variables aleatorias independientes, ambas con distribución exponencial de parámetro λ . a) Encuentre las funciones de distribución condicionales de $\min(X, Y)$ y $\max(X, Y)$ dado que $Y = y$, para $y > 0$. b) Utilice los resultados de la parte a para calcular $E[\min(X, Y) | Y]$ y $E[\max(X, Y) | Y]$.

EJERCICIO 4.57. Utilice la conclusión del ejemplo 4.41 para interpretar el resultado del ejercicio 2.19.

EJERCICIO 4.58. Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que las variables aleatorias $Y_1 = T_1, Y_2 = T_2 - T_1, Y_3 = T_3 - T_2, \dots$ son independientes y, para $t > 0$ y $n \in \mathbb{N}$, la distribución conjunta de T_1, \dots, T_n , dado que $T_{n+1} = t$, es la misma que la de los estadísticos de orden correspondientes a n variables aleatorias independientes, todas con distribución uniforme en el intervalo $(0, t)$. Asumiendo que T_1, T_2, \dots son absolutamente continuas y que sus funciones de densidad son diferenciables, demuestre que Y_1, Y_2, \dots tienen distribución exponencial con parámetro común.

EJERCICIO 4.59. Sea Y una variable aleatoria con distribución Poisson de parámetro λ y supongamos que para cada valor n de Y , X es una variable aleatoria con distribución binomial de parámetros n y p . Encuentre a) la distribución de X y b) la distribución condicional de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$.

EJERCICIO 4.60. Supongamos que se envía una señal aleatoria X desde un lugar A de tal manera que su distribución es $N(\mu, \sigma^2)$. Supongamos además que, cuando $X = x$, el valor Y , que se recibe en un lugar B , tiene distribución $N(x, a^2)$, en donde a es una constante distinta de cero. a) Encuentre $E(Y)$, $\text{Var}(Y)$ y $\text{Cov}(X, Y)$. b) Demuestre que la distribución conjunta de la pareja X, Y es normal bivariada. c) Encuentre una función de densidad de Y . d) Dado que $Y = y$, ¿cuál es el mejor estimador de X en el sentido de la media cuadrática?

EJERCICIO 4.61. Sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y supongamos que, para cada valor x de X , Y es una variable aleatoria con distribución normal de parámetros $\mu = \alpha + \beta x$ y σ^2 , en donde α, β y σ son constantes. Encuentre $E(Y)$, $\text{Var}(Y)$ y una función de densidad conjunta de la pareja X, Y .

EJERCICIO 4.62. *Un minero está atrapado en una mina que tiene 3 túneles. El primero lo lleva a la salida después de media hora, el segundo lo regresa al mismo lugar después de 1 hora y el tercero también lo regresa al mismo lugar después de 2 horas. Supongamos que, en cada ocasión, el minero escoge el primero, segundo y tercer túnel con probabilidades 0.2, 0.5 y 0.3, respectivamente. ¿Cuál es la esperanza del tiempo que lleva al minero salir de la mina?*

EJERCICIO 4.63. *Una rata de laboratorio está encerrada en un lugar que contiene N salidas. Para $1 \leq k \leq N - 1$, la salida número k conduce a la rata al mismo lugar después de k minutos, mientras que la salida número N lleva a un camino que conduce a la rata a la verdadera salida después N minutos. Supongamos que la rata escoge siempre al azar una de las N salidas. ¿Cuál es el número esperado de minutos que le lleva a la rata llegar a la verdadera salida?*

EJERCICIO 4.64. *Una urna contiene bolas rojas y negras de tal manera que la proporción de bolas rojas que contiene es igual a p . Se van seleccionando bolas de la urna al azar, una a una y con reemplazo, hasta que se obtienen r bolas rojas en forma consecutiva. Si X es el número de bolas que se seleccionan hasta que se detiene el proceso, encuentre $E[X]$.*

EJERCICIO 4.65. *Sea X_1 un número que se elige al azar en el intervalo $(0, 1)$, X_2 un número que se elige al azar en el intervalo $(0, X_1)$, X_3 un número que se elige al azar en el intervalo $(0, X_2)$, etc. Encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.*

EJERCICIO 4.66. *Una urna contiene una bola roja y una bola negra. Se elige al azar una bola de la urna y se reemplaza agregando una bola del mismo color que la seleccionada. Si este proceso se repite indefinidamente y llamamos X_n a la proporción de bolas rojas en la urna después de la n -sima elección, encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.*

EJERCICIO 4.67. *Una urna contiene inicialmente r bolas Rojas y s bolas Negras. Se agregan a la urna 1 bola Roja y 2 Negras e inmediatamente después se seleccionan, al azar y sin reemplazo, 3 bolas de la misma. Supongamos que este proceso se repite indefinidamente y llamemos X_n al número de bolas Rojas que quedan en la urna después del paso n . Encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.*

EJERCICIO 4.68. *Una persona está jugando un juego de azar en el cual gana con probabilidad p , de tal manera que $p > \frac{1}{2}$. La estrategia que sigue consiste en apostar en cada juego la fracción $2p - 1$ de su fortuna en ese momento. Supongamos que la fortuna inicial del jugador es x y llamemos X_n a su fortuna después de n juegos. Encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.*

EJERCICIO 4.69. *Una urna contiene inicialmente a bolas azules y r bolas rojas. Se selecciona al azar una bola de la urna; si es roja, se regresa, si no, se reemplaza por una roja. Supongamos que este proceso se repite indefinidamente y llamemos X_n al*

número de bolas rojas que quedan en la urna después del paso n . Encuentre $E[X_n]$ para cualquier $n \in \mathbb{N}$.

EJERCICIO 4.70. Una moneda se elige al azar de una colección de monedas, de tal manera que la probabilidad p de obtener cara puede considerarse como seleccionada al azar en el intervalo $(0, 1)$. Si la moneda se lanza dos veces en forma consecutiva, encuentre la probabilidad de que a) en el primer lanzamiento se obtenga cara y b) se obtenga cara en ambos lanzamientos.

EJERCICIO 4.71. Supongamos que el número esperado de accidentes por semana que hay en una fábrica es igual a 5. Supongamos también que el número de trabajadores afectados en un accidente particular es una variable aleatoria, independiente del número de accidentes, de esperanza 2.5. Encuentre el número esperado de trabajadores afectados por algún accidente en una semana. Argumente claramente su respuesta utilizando distribuciones condicionales. ¿Se podría asegurar la misma respuesta si el número de trabajadores afectados en un accidente particular no fuera independiente del número de accidentes? En caso afirmativo, demuéstrello; en caso contrario, exhiba un contraejemplo.

EJERCICIO 4.72. Sea X un número que se elige al azar en el intervalo $(0, 1)$ y Y un número que se elige al azar en el intervalo $(0, X)$. Encuentre una función de densidad Y .

EJERCICIO 4.73. Supongamos que un cierto evento ocurre en los tiempos aleatorios T_1, T_2, \dots , de tal manera que si, para $t \geq 0$, X_t es el número de veces que ocurre el evento hasta el tiempo t , entonces la familia de variables aleatorias $\{X_t\}_{t \geq 0}$ forma un proceso de Poisson de parámetro λ . Encuentre la distribución del número de eventos que ocurren en el intervalo de tiempo $[0, T]$, en donde a) T es una variable aleatoria con distribución exponencial de parámetro ν . b) T es una variable aleatoria con distribución uniforme en el intervalo $[0, a]$.

EJERCICIO 4.74. Sea Y una variable aleatoria con distribución beta de parámetros α y β y supongamos que para cada valor y de Y , X es una variable aleatoria con distribución binomial de parámetros n y $p = y$. Encuentre a) la distribución de X , b) $E[X]$ y $\text{Var}(X)$ y c) la distribución condicional de Y dado que $X = x$, para $x \in \{0, \dots, n\}$.

Nota: La distribución que se obtiene para X es conocida como la distribución de Polya.

EJERCICIO 4.75. Sea Y una variable aleatoria con uniforme en el intervalo $(0, 1)$ y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución geométrica de parámetro y . a) Encuentre $E[X]$. b) Para $x \in \{0, 1, \dots\}$, encuentre e identifique la distribución condicional de Y dado que $X = x$. c) Encuentre $E[(X + Y)^2 | X]$.

EJERCICIO 4.76. Sea Y una variable aleatoria con distribución beta de parámetros α y β y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución geométrica de parámetro y . Encuentre: a) la distribución de X , b) $E[X]$, c) la distribución de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$ y d) $E[Y | X]$.

EJERCICIO 4.77. Sea Y una variable aleatoria con distribución beta de parámetros α y β y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución binomial negativa de parámetros r y y . a) Encuentre $E[X]$. b) Encuentre e identifique la distribución condicional de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$. c) Para $\alpha = 1$, $\beta = 1$ y $r = 2$, encuentre $E[(X + Y)^2 | X]$.

EJERCICIO 4.78. Sea Y una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y supongamos que para cada valor y de Y , X es una variable aleatoria con distribución geométrica de parámetro y . Encuentre: a) la distribución de X , b) $E[X]$, c) la distribución de Y dado que $X = k$, para $k \in \{0, 1, \dots\}$ y d) $E[Y | X]$.

EJERCICIO 4.79. Sea Y una variable aleatoria con distribución gama de parámetros α y λ y supongamos que, para cada valor y de Y , X es una variable aleatoria con distribución Poisson de parámetro y . Encuentre a) la distribución de X , b) $E[X]$, c) la distribución condicional de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$ y d) $E[Y | X]$.

EJERCICIO 4.80. Considerando la misma situación que en el ejemplo 4.70, a) demuestre que dado que la persona seleccionada tuvo j accidentes en el último año, la distribución del número de accidentes en el presente año tiene distribución binomial negativa. ¿Cuál es la probabilidad de que una persona, seleccionada al azar, tenga exactamente 2 accidentes en un año dado que b) en el año anterior no tuvo accidentes, c) en el año anterior tuvo 1 accidente.

EJERCICIO 4.81. Cada una de N bolas se coloca al azar en alguna de r cajas, en donde N es una variable aleatoria con distribución Poisson de parámetro λ . Demuestre que el número de cajas que quedan vacías tiene distribución binomial.

Parte 2

CONVERGENCIA

CAPÍTULO 5

TEOREMAS LÍMITE

Todo fluye, nada permanece ni persiste nunca lo mismo.

Heráclito

Se sabe que el Cálculo de Probabilidades tiene como soporte esencialmente un único teorema, la ley de los grandes números. Se puede decir que la teoría tiene como único objetivo el demostrar ese teorema y algunos otros que se le relacionan.

Paul Pierre Lévy

El surgimiento del Cálculo de Probabilidades, como disciplina matemática independiente, tiene como base las soluciones que, durante el periodo que va del año 1654 al año 1657, dieron Blaise Pascal, Pierre de Fermat y Christiaan Huygens a varios problemas, los cuales se analizan en el capítulo refulcalpro de este volumen. Pero, si bien los trabajos de Pascal, Fermat y Huygens permitieron el desarrollo de métodos generales para resolver problemas de probabilidad, éstos se limitaban a un tipo muy particular, relacionados con juegos de azar, los cuales no eran suficientes para darle un lugar dentro de las matemáticas a la Teoría de la Probabilidad.

El gran impulso para el desarrollo de una Teoría de la Probabilidad, que le haría ganar un lugar dentro de las matemáticas, proviene de los llamados teoremas límite, los cuales se refieren al comportamiento a largo plazo de sucesiones de variables aleatorias. El primero de estos resultados, que para algunos autores marca verdaderamente el inicio de la historia de la Teoría de la Probabilidad, se debe a Jacques Bernoulli, quien dedicó 20 años de su vida a la búsqueda de una prueba matemática de la relación que existe entre la probabilidad de un evento y la frecuencia relativa con la que éste

ocurre en una serie grande de repeticiones del correspondiente experimento aleatorio. El resultado, conocido como teorema de Bernoulli, se publicó en el año 1718, cinco años después de la muerte de su autor.

Puede decirse que, a partir de la publicación del teorema de Bernoulli, el motor de desarrollo de la Teoría de la Probabilidad fue la búsqueda de resultados que permitieran mejorar y generalizar ese teorema. Vendrían después los teoremas de de Moivre y de Poisson, relativos a la aproximación de una distribución binomial mediante una distribución normal y una distribución Poisson, respectivamente, los cuales fueron publicados en los años 1730 y 1800, respectivamente.

Este proceso continuaría desarrollándose y recibiría un gran impulso, entre 1870 y 1900, con los trabajos de la llamada escuela rusa, representada por Pafnuty Lvovich Chebyshev, Andrei Andreyevich Markov y Aleksandr Mikhailovich Lyapunov, entre otros, los cuales conducirían a la forma general que se dio a los teoremas límite, entre 1900 y 1930, con la formulación de las leyes de los grandes números y el teorema del límite central, tanto en su forma clásica, relativa a la convergencia a la distribución normal, como en su forma moderna, relativa a la convergencia a cualquier otro tipo de distribución, sobresaliendo en este periodo los trabajos de Aleksandr Yakovlevich Khintchine, Andrey Nikolaevich Kolmogorov, J. W. Lindeberg, William Feller y Paul Pierre Lévy, entre otros.

Como puede verse, fueron más de 200 años de historia de la Teoría de la Probabilidad guiada por el estudio de los teoremas límite.

5.1. Diferentes tipos de convergencia

DEFINICIÓN 5.1 (Convergencia en probabilidad). *Se dice que una sucesión $\{X_n\}$ de variables aleatorias converge en probabilidad a la variable aleatoria X si:*

$$\lim_{n \rightarrow \infty} P[|X_n - X| > \varepsilon] = 0$$

para cualquier $\varepsilon > 0$. En este caso se escribirá $X_n \xrightarrow{P} X$.

Obviamente si una sucesión $\{X_n\}$ converge en probabilidad a X , entonces cualquier subsucesión de $\{X_n\}$ también converge en probabilidad a X .

PROPOSICIÓN 5.2. *Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{P} X$ y $X_n \xrightarrow{P} Y$, entonces $P[X = Y] = 1$.*

Demostración

Como $|X - Y| \leq |X_n - X| + |X_n - Y|$, entonces:

$$[|X - Y| > \varepsilon] \subset [|X_n - X| + |X_n - Y| > \varepsilon]$$

Además, para cualquier $\varepsilon > 0$, se tiene:

$$[|X_n - X| + |X_n - Y| > \varepsilon] \subset [|X_n - X| > \frac{\varepsilon}{2}] \cup [|X_n - Y| > \frac{\varepsilon}{2}]$$

Por lo tanto:

$$P[|X - Y| > \varepsilon] \leq P[|X_n - X| > \frac{\varepsilon}{2}] + P[|X_n - Y| > \frac{\varepsilon}{2}]$$

Así que, tomando límites, se obtiene $P[|X - Y| > \varepsilon] = 0$ para cualquier $\varepsilon > 0$.

Finalmente, $[|X - Y| > 0] = \bigcup_{n=1}^{\infty} [|X - Y| > \frac{1}{n}]$, así que:

$$P[|X - Y| > 0] \leq \sum_{n=1}^{\infty} P[|X - Y| > \frac{1}{n}] = 0$$

■

PROPOSICIÓN 5.3. *Sea c una constante y $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{P} X$, entonces $cX_n \xrightarrow{P} cX$.*

Demostración

$$\lim_{n \rightarrow \infty} P[|cX_n - cX| > \varepsilon] = \lim_{n \rightarrow \infty} P\left[|X_n - X| > \frac{\varepsilon}{|c|}\right] = 0$$

■

PROPOSICIÓN 5.4. *Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$, entonces $X_n + Y_n \xrightarrow{P} X + Y$.*

Demostración

Como $|X_n - X + Y_n - Y| \leq |X_n - X| + |Y_n - Y|$, se tiene:

$$[|X_n - X + Y_n - Y| > \varepsilon] \subset [|X_n - X| > \frac{\varepsilon}{2}] \cup [|Y_n - Y| > \frac{\varepsilon}{2}]$$

Así que:

$$\begin{aligned} & \lim_{n \rightarrow \infty} P[|X_n + Y_n - X - Y| > \varepsilon] \\ & \leq \lim_{n \rightarrow \infty} P[|X_n - X| > \frac{\varepsilon}{2}] + \lim_{n \rightarrow \infty} P[|Y_n - Y| > \frac{\varepsilon}{2}] = 0 \end{aligned}$$

■

PROPOSICIÓN 5.5. *Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{P} X$, entonces $X_n^2 \xrightarrow{P} X^2$.*

Demostación

Como $\sum_{k=0}^{\infty} P[k \leq |X| < k+1] = 1$, entonces dada $\delta > 0$ existe M tal que:

$$P[|X| > M] \leq P[|X| \geq M] = \sum_{k=M}^{\infty} P[k \leq |X| < k+1] < \frac{\delta}{2}$$

También, como $X_n \xrightarrow{P} X$, existe N tal que, si $n \geq N$, entonces:

$$P[|X_n - X| > M] < \frac{\delta}{2}$$

Además, $|X_n + X| \leq |X_n - X| + 2|X|$, así que:

$$[|X_n + X| > 4M] \subset [|X_n - X| > 2M] \cup [|X| > M]$$

Por lo tanto, para $n \geq N$, se tiene:

$$P[|X_n + X| > 4M] \leq P[|X_n - X| > 2M] + P[|X| > M] < \delta$$

Así que, dada $\varepsilon > 0$:

$$\begin{aligned} P[|X_n^2 - X^2| > \varepsilon] &= P[|(X_n + X)(X_n - X)| > \varepsilon] \\ &= P[|(X_n + X)(X_n - X)| > \varepsilon, |X_n + X| \leq 4M] \\ &\quad + P[|(X_n + X)(X_n - X)| > \varepsilon, |X_n + X| > 4M] \\ &\leq P[|(X_n + X)(X_n - X)| > \varepsilon, 0 < |X_n + X| \leq 4M] \\ &\quad + P[|(X_n + X)(X_n - X)| > \varepsilon, |X_n + X| > 4M] \\ &= P[|X_n - X| > \frac{\varepsilon}{4M}, 0 < |X_n + X| \leq 4M] \\ &\quad + P[|(X_n + X)(X_n - X)| > \varepsilon, |X_n + X| > 4M] \\ &\leq P[|X_n - X| > \frac{\varepsilon}{4M}] + P[|X_n + X| > 4M] \\ &< P[|X_n - X| > \frac{\varepsilon}{4M}] + \delta \end{aligned}$$

De manera que tomando límites, se obtiene:

$$\limsup_{n \rightarrow \infty} P[|X_n^2 - X^2| > \varepsilon] \leq \delta \text{ para cualquier } \delta > 0$$

$$\text{Por lo tanto, } \lim_{n \rightarrow \infty} P[|X_n^2 - X^2| > \varepsilon] = 0.$$

■

COROLARIO 5.6. Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$, entonces $X_n Y_n \xrightarrow{P} XY$.

Demostración

Como $X_n Y_n = \frac{1}{4} [(X_n + Y_n)^2 - (X_n - Y_n)^2]$, entonces:

$$X_n Y_n \xrightarrow{P} \frac{1}{4} [(X + Y)^2 - (X - Y)^2] = XY$$

■

EJEMPLO 5.7. Sea $\Omega = (0, 1]$ y P la medida de Lebesgue sobre Ω , es decir, la medida de probabilidad sobre el intervalo $(0, 1]$ que asigna a cada intervalo su longitud. Para cada $n \in \mathbb{N}$, definamos $X_n = I_{(0, \frac{1}{n})}$, es decir:

$$X_n(\omega) = \begin{cases} 1 & \text{si } \omega < \frac{1}{n} \\ 0 & \text{en otro caso} \end{cases}$$

Dada $\varepsilon > 0$, se tiene:

$$P[|X_n| > \varepsilon] = P[X_n > \varepsilon] = \begin{cases} \frac{1}{n} & \text{si } \varepsilon < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Así que, en cualquier caso, $\lim_{n \rightarrow \infty} P[|X_n| > \varepsilon] = 0$. Por lo tanto, $X_n \xrightarrow{P} 0$.

EJEMPLO 5.8. Sea $\Omega = (0, 1]$ y P la medida de Lebesgue sobre Ω . Para cada $n \in \mathbb{N}$, definamos:

$$X_n = \begin{cases} I_{(0, \frac{1}{n}]} & \text{si } n \text{ es impar} \\ I_{(\frac{1}{n}, 1]} & \text{si } n \text{ es par} \end{cases}$$

Dada $\varepsilon > 0$ y n impar, se tiene:

$$P[|X_n| > \varepsilon] = P[X_n > \varepsilon] = \begin{cases} \frac{1}{n} & \text{si } \varepsilon < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Así que $X_{2n+1} \xrightarrow{P} 0$.

Por otro lado, dada $\varepsilon > 0$ y n par, se tiene:

$$P[|X_n - 1| > \varepsilon] = \begin{cases} \frac{1}{n} & \text{si } \varepsilon < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Así que $X_{2n} \xrightarrow{P} 1$.

Por lo tanto, la sucesión $\{X_n\}$ no converge en probabilidad.

▲

En los dos ejemplos anteriores, la convergencia o no convergencia en probabilidad coincide con la convergencia o no convergencia de las variables aleatorias X_n vistas como funciones. Sin embargo éste no es siempre el caso. La convergencia en probabilidad significa que es pequeña la probabilidad de que $|X_n - X|$ sea grande, pero la sucesión de funciones X_n pudiera ni siquiera ser convergente, como se muestra en el siguiente ejemplo:

EJEMPLO 5.9. Sea $\Omega = (0, 1]$ y P la medida de Lebesgue sobre Ω . Para $i \in \mathbb{N}$ y $j \in \{1, \dots, i\}$, definamos $Y_{i,j} = I_{(\frac{i-1}{i}, \frac{i}{i}]}$ y ordenemos la familia de variables aleatorias $\{Y_{i,j}\}$, primero de acuerdo al primer subíndice i y después, fijando el subíndice i , de acuerdo al subíndice j . De esta forma, se obtiene la sucesión $X_1 = Y_{11}$, $X_2 = Y_{21}$, $X_3 = Y_{22}$, $X_4 = Y_{31}$, $X_5 = Y_{32}$, $X_6 = Y_{33}$, \dots . En general, si $\frac{i(i-1)}{2} < n \leq \frac{i(i+1)}{2}$ y $n = \frac{i(i-1)}{2} + j$, entonces $X_n = Y_{ij}$.

Evidentemente, la sucesión de funciones $\{X_n\}$ no converge para ningún $\omega \in (0, 1]$, sin embargo, $P[X_{ij} = 1] = \frac{1}{i}$ y $P[X_{ij} = 0] = 1 - \frac{1}{i}$, así que $\lim_{n \rightarrow \infty} P[|X_n| > \varepsilon] = 0$ para cualquier $\varepsilon > 0$. Por lo tanto, $X_n \xrightarrow{P} 0$.

DEFINICIÓN 5.10 (Convergencia en distribución). Se dice que una sucesión $\{X_n\}$ de variables aleatorias converge en distribución a la variable aleatoria X si:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

para cualquier número real x en el cual F_X es continua. En este caso se escribirá $X_n \xrightarrow{D} X$.

Obviamente si una sucesión $\{X_n\}$ converge en distribución a X , entonces cualquier subsucesión de $\{X_n\}$ también converge en distribución a X .

PROPOSICIÓN 5.11. Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{D} X$ y $X_n \xrightarrow{D} Y$, entonces $F_X = F_Y$.

Demostración

De la definición de convergencia en distribución se sigue inmediatamente que $F_X(z) = F_Y(z)$ para cualquier número real z tal que F_X y F_Y son continuas en z . Pero como el conjunto de discontinuidades de F_X y de F_Y es a lo más numerable, entonces el conjunto de números reales z para los cuales tanto F_X como F_Y son continuas en z es denso en \mathbb{R} . El resultado se sigue entonces de la continuidad por la derecha de F_X y F_Y . ■

También se tienen los siguientes dos resultados, cuya demostración se deja como ejercicio.

PROPOSICIÓN 5.12. *Sea c una constante y $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{D} X$, entonces $cX_n \xrightarrow{D} cX$.*

PROPOSICIÓN 5.13. *Sea c una constante y $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{D} X$, entonces $X_n + c \xrightarrow{D} X + c$.*

EJEMPLO 5.14. *Sea $\{X_n\}$ la sucesión del ejemplo 5.7, entonces $X_n \xrightarrow{D} 0$.*

EJEMPLO 5.15. *Sea $\{X_n\}$ la sucesión del ejemplo 5.8, entonces $\{X_n\}$ no converge en distribución.*

EJEMPLO 5.16. *Sea $\{X_n\}$ la sucesión del ejemplo 5.9, entonces $X_n \xrightarrow{D} 0$.*

EJEMPLO 5.17. *El teorema de de Moivre Laplace constituye un ejemplo básico de convergencia en distribución. Este resultado establece, en particular, que si $a \in \mathbb{R}$ y, para cada $n \in \mathbb{R}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p \in (0, 1)$. Entonces:*

$$\lim_{n \rightarrow \infty} P \left[\frac{X_n - np}{\sqrt{npq}} > a \right] = \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-\frac{1}{2}x^2} dx$$

Es decir, $X_n \xrightarrow{D} X$, en donde X es una variable aleatoria con distribución normal estándar.

EJEMPLO 5.18. *El teorema de Poisson constituye otro ejemplo básico de convergencia en distribución. Este resultado establece que si, para cada $n \in \mathbb{R}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p \in (0, 1)$ de tal manera que $\lambda = np$ es constante, entonces, para cualquier $k \in \{0, 1, \dots\}$, se tiene:*

$$\lim_{n \rightarrow \infty} P [X_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Por lo tanto, para cualquier $x \geq 0$, se tiene:

$$\lim_{n \rightarrow \infty} P [X_n \leq x] = \sum_{k=0}^{[x]} \frac{\lambda^k e^{-\lambda}}{k!}$$

en donde $[x]$ denota a la parte entera de x .

Es decir, $X_n \xrightarrow{D} X$, en donde X es una variable aleatoria con distribución Poisson de parámetro λ .

DEFINICIÓN 5.19 (**Convergencia casi segura**). *Se dice que una sucesión $\{X_n\}$ de variables aleatorias converge casi seguramente a la variable aleatoria X si:*

$$P[\lim_{n \rightarrow \infty} X_n = X] = 1$$

En este caso se escribirá $X_n \xrightarrow{c.s.} X$.

Obviamente si una sucesión $\{X_n\}$ converge casi seguramente a X , entonces cualquier subsucesión de $\{X_n\}$ también converge casi seguramente a X .

La demostración de la siguiente proposición es inmediata y las tres que le siguen se siguen inmediatamente de los resultados análogos para sucesiones de números reales.

PROPOSICIÓN 5.20. *Sea $\{X_n\}$ una sucesión de variables aleatorias tales que $X_n \xrightarrow{c.s.} X$ y $X_n \xrightarrow{c.s.} Y$, entonces $P[X = Y] = 1$.*

PROPOSICIÓN 5.21. *Sea c una constante y $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{c.s.} X$, entonces $cX_n \xrightarrow{c.s.} cX$.*

PROPOSICIÓN 5.22. *Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$, entonces $X_n + Y_n \xrightarrow{c.s.} X + Y$.*

PROPOSICIÓN 5.23. *Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{c.s.} X$ y $Y_n \xrightarrow{c.s.} Y$, entonces $X_n Y_n \xrightarrow{c.s.} XY$.*

EJEMPLO 5.24. *Sea $\{X_n\}$ la sucesión del ejemplo 5.7, entonces $X_n \xrightarrow{c.s.} 0$.*

EJEMPLO 5.25. *Sea $\{X_n\}$ la sucesión del ejemplo 5.8, entonces $\{X_n\}$ no converge casi seguramente.*

EJEMPLO 5.26. *Sea $\{X_n\}$ la sucesión del ejemplo 5.9, entonces $\{X_n\}$ no converge casi seguramente.*

5.2. Relación entre modos de convergencia

PROPOSICIÓN 5.27. *Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{P} X$, entonces $X_n \xrightarrow{D} X$.*

Demostración

Para $\varepsilon > 0$, $n \in \mathbb{N}$ y $t \in \mathbb{R}$, se tiene:

$$\begin{aligned} F_X(t - \varepsilon) &= P[X \leq t - \varepsilon] \\ &= P[X \leq t - \varepsilon, |X - X_n| > \varepsilon] + P[X \leq t - \varepsilon, |X - X_n| \leq \varepsilon] \\ &= P[X \leq t - \varepsilon, |X - X_n| > \varepsilon] + P[X \leq t - \varepsilon, X - \varepsilon \leq X_n \leq X + \varepsilon] \end{aligned}$$

$$\begin{aligned}
&\leq P[|X - X_n| > \varepsilon] + P[X_n \leq t] = P[|X - X_n| > \varepsilon] + F_{X_n}(t) \\
F_{X_n}(t) &= P[X_n \leq t] = P[X_n \leq t, |X - X_n| > \varepsilon] + P[X_n \leq t, |X - X_n| \leq \varepsilon] \\
&= P[X_n \leq t, |X - X_n| > \varepsilon] + P[X_n \leq t, X_n - \varepsilon \leq X \leq X_n + \varepsilon] \\
&\leq P[|X - X_n| > \varepsilon] + P[X \leq t + \varepsilon] = P[|X - X_n| > \varepsilon] + F_X(t + \varepsilon)
\end{aligned}$$

Así que, para cualquier $\varepsilon > 0$, $n \in \mathbb{N}$ y $t \in \mathbb{R}$, se tiene:

$$F_X(t - \varepsilon) - P[|X - X_n| > \varepsilon] \leq F_{X_n}(t) \leq F_X(t + \varepsilon) + P[|X - X_n| > \varepsilon]$$

Tomando límites cuando $n \rightsquigarrow \infty$ y utilizando el hecho de que $X_n \xrightarrow{P} X$, se obtiene:

$$F_X(t - \varepsilon) \leq \liminf_{n \rightsquigarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightsquigarrow \infty} F_{X_n}(t) \leq F_X(t + \varepsilon)$$

Ahora, si t es un punto de continuidad de F_X , entonces, tomando límites cuando $\varepsilon \rightsquigarrow 0$, se obtiene:

$$F_X(t) \leq \liminf_{n \rightsquigarrow \infty} F_{X_n}(t) \leq \limsup_{n \rightsquigarrow \infty} F_{X_n}(t) \leq F_X(t)$$

Así que $\lim_{n \rightsquigarrow \infty} F_{X_n}(t) = F_X(t)$.

■

El ejemplo siguiente muestra que el inverso de la proposición 5.27 no es válido en general.

EJEMPLO 5.28. *Sea X una variable aleatoria con distribución normal estándar y, para cada $n \in \mathbb{N}$, definamos:*

$$X_n = \begin{cases} X & \text{si } n \text{ es impar} \\ -X & \text{si } n \text{ es par} \end{cases}$$

Entonces $F_{X_n}(x) = F_X(x)$ para cualquier $x \in \mathbb{R}$, así que $X_n \xrightarrow{D} X$. Sin embargo, $|X_{2n} - X| = 2|X|$ para cualquier $n \in \mathbb{N}$, así que $P[|X_{2n} - X| > \varepsilon] = P[|X| > \frac{\varepsilon}{2}]$ para cualquier $\varepsilon > 0$. Por lo tanto $\lim_{n \rightsquigarrow \infty} P[|X_{2n} - X| > \varepsilon] = P[|X| > \frac{\varepsilon}{2}] > 0$, así que la sucesión $\{X_n\}$ no converge a X en probabilidad. De hecho, la sucesión $\{X_n\}$ no converge en probabilidad a ninguna variable aleatoria pues $X_{2n+1} \rightsquigarrow X$ en probabilidad, así que si $X_n \rightsquigarrow Y$ en probabilidad, entonces $P[X = Y] = 1$, así que se debería tener $X_n \rightsquigarrow X$ en probabilidad, lo cual es falso.

▲

Se tiene el siguiente resultado parcial:

PROPOSICIÓN 5.29. *Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{D} 0$, entonces $X_n \xrightarrow{P} 0$.*

Demostración

La hipótesis nos dice que:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x > 0 \end{cases}$$

Además, para $\varepsilon > 0$:

$$\begin{aligned} P[|X_n| > \varepsilon] &= P[X_n > \varepsilon] + P[X_n < -\varepsilon] \leq P[X_n > \varepsilon] + P[X_n \leq -\varepsilon] \\ &= 1 - F_{X_n}(\varepsilon) + F_{X_n}(-\varepsilon) \end{aligned}$$

Así que, $\lim_{n \rightarrow \infty} P[|X_n| > \varepsilon] = 0$. ■

COROLARIO 5.30. *Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que:*

$$X_n \xrightarrow{D} 0 \text{ y } Y_n \xrightarrow{D} 0, \text{ entonces } X_n + Y_n \xrightarrow{D} 0.$$

COROLARIO 5.31. *Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{D} 0$ y $Y_n \xrightarrow{D} 0$, entonces $X_n Y_n \xrightarrow{D} 0$.*

PROPOSICIÓN 5.32. *Sea $\{X_n\}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{c.s.} X$, entonces $X_n \xrightarrow{P} X$.*

Demostración

Como $P[\lim_{n \rightarrow \infty} X_n = 0] = 1$, existe un conjunto $\Omega_0 \subset \Omega$ de probabilidad 0 tal que si $\omega \in \Omega_0^c$ entonces $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Así que, dado $\omega \in \Omega_0^c$ y $\varepsilon > 0$, existe N tal que $|X_n(\omega)| \leq \varepsilon$ para cualquier $n \geq N$, esto significa que $\omega \in \bigcap_{n=N}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}$.

Dicho de otra forma, si $\omega \in \Omega_0^c$, entonces, dada cualquier $\varepsilon > 0$:

$$\omega \in \bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}$$

para alguna m , lo cual a su vez significa que $\omega \in \bigcup_{m=1}^{\infty} [\bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}]$. Así que:

$$\Omega_0^c \subset \bigcup_{m=1}^{\infty} [\bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}]$$

Por lo tanto:

$$P\left(\bigcup_{m=1}^{\infty} \left[\bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}\right]\right) = 1$$

Sea $B_m(\varepsilon) = \bigcap_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| \leq \varepsilon\}$, entonces la sucesión de eventos $B_1(\varepsilon), B_2(\varepsilon), \dots$ es monótona creciente, así que:

$$\lim_{m \rightarrow \infty} P[B_m(\varepsilon)] = P\left(\bigcup_{m=1}^{\infty} B_m(\varepsilon)\right) = 1$$

De lo cual se sigue, $\lim_{m \rightarrow \infty} P[B_m^c(\varepsilon)] = 0$.

Pero $B_m^c(\varepsilon) = \bigcup_{n=m}^{\infty} \{\omega \in \Omega : |X_n(\omega)| > \varepsilon\}$, así que $\{|X_m| > \varepsilon\} \subset B_m^c(\varepsilon)$. Por lo tanto:

$$\lim_{m \rightarrow \infty} P[|X_m| > \varepsilon] \leq \lim_{m \rightarrow \infty} P[B_m^c(\varepsilon)] = 0.$$

■

El inverso de la proposición 5.32 no es válido en general. Para un ejemplo, considérese la sucesión $\{X_n\}$ del ejemplo 5.9, la cual converge en probabilidad, pero no converge casi seguramente.

5.3. Lema de Borel-Cantelli y convergencia casi segura

PROPOSICIÓN 5.33. *Sea $\{X_n\}$ una sucesión de variables aleatorias, entonces $X_n \xrightarrow{c.s.} 0$ si y sólo si:*

$$P[\{\omega \in \Omega : |X_n(\omega)| > \varepsilon \text{ para una infinidad de valores de } n\}] = 0$$

para cualquier $\varepsilon > 0$.

Demostración

Supongamos primero que $X_n \xrightarrow{c.s.} 0$ y, para cada $\varepsilon > 0$, sea:

$$A(\varepsilon) = \{\omega \in \Omega : |X_n(\omega)| > \varepsilon \text{ para una infinidad de valores de } n\}$$

Como $P[\lim_{n \rightarrow \infty} X_n = 0] = 1$, existe un conjunto $\Omega_0 \subset \Omega$ de probabilidad 1 tal que si $\omega \in \Omega_0$ entonces $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Así que, dado $\omega \in \Omega_0$ y $\varepsilon > 0$, existe N tal que $|X_n(\omega)| < \varepsilon$ para cualquier $n \geq N$. Por lo tanto, si $\omega \in A(\varepsilon)$, entonces $\omega \in \Omega_0^c$, así que $P[A(\varepsilon)] \leq P[\Omega_0^c] = 0$.

Inversamente, supongamos que:

$$P[\{\omega \in \Omega : |X_n(\omega)| > \varepsilon \text{ para una infinidad de valores de } n\}] = 0$$

para cualquier $\varepsilon > 0$

Para cada $r \in \mathbb{N}$, sea:

$$B_r = \left\{ \omega \in \Omega : |X_n(\omega)| > \frac{1}{r} \text{ para una infinidad de valores de } n \right\}$$

Se tiene $P(B_r) = 0$ para cualquier $r \in \mathbb{N}$ y la sucesión de eventos B_1^c, B_2^c, \dots es monótona decreciente, así que:

$$P\left(\bigcap_{r=1}^{\infty} B_r^c\right) = \lim_{r \rightarrow \infty} P(B_r^c) = 1$$

Pero, $B_r^c = \left\{ \omega \in \Omega : \text{Existe } N(\omega) \text{ tal que } |X_n(\omega)| \leq \frac{1}{r} \text{ para cualquier } n \geq N(\omega) \right\}$. De manera que si $\omega \in \bigcap_{r=1}^{\infty} B_r^c$, entonces para cualquier $r \in \mathbb{N}$ existe $N(\omega)$ tal que $|X_n(\omega)| \leq \frac{1}{r}$ para cualquier $n \geq N(\omega)$. En particular, dada $\varepsilon > 0$ sea $r \in \mathbb{N}$ tal que $\frac{1}{r} < \varepsilon$ y $N(\omega)$ tal que $|X_n(\omega)| \leq \frac{1}{r}$ para cualquier $n \geq N(\omega)$, entonces $|X_n(\omega)| < \varepsilon$ para cualquier $n \geq N(\omega)$, lo cual significa que $\lim_{n \rightarrow \infty} X_n(\omega) = 0$. Es decir, $\bigcap_{r=1}^{\infty} B_r^c \subset [\lim_{n \rightarrow \infty} X_n = 0]$ y entonces $P[\lim_{n \rightarrow \infty} X_n = 0] \geq P\left(\bigcap_{r=1}^{\infty} B_r^c\right) = 1$.

COROLARIO 5.34. *Sea $\{X_n\}$ una sucesión de variables aleatorias y X otra variable aleatoria, entonces $X_n \xrightarrow{c.s.} X$ si y sólo si:*

$$P\left[\left\{ \omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \text{ para una infinidad de valores de } n \right\}\right] = 0$$

para cualquier $\varepsilon > 0$.

PROPOSICIÓN 5.35 (Lema de Borel-Cantelli). *Sea A_1, A_2, \dots una sucesión de eventos tales que $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces:*

$$P\left[\left\{ \omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n \right\}\right] = 0$$

Demostración

Sea $A = \left\{ \omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n \right\}$.

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de eventos B_m es monótona decreciente y $A = \bigcap_{m=1}^{\infty} B_m$, así que:

$$P(A) = P\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m \rightarrow \infty} P\left[\bigcup_{n=m}^{\infty} A_n\right] \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} P(A_n) = 0$$

■

COROLARIO 5.36. *Sea X_1, X_2, \dots una sucesión de variables aleatorias tales que:*

$$\sum_{n=1}^{\infty} P[|X_n| > \varepsilon] < \infty$$

para cualquier $\varepsilon > 0$.

Entonces:

$$X_n \xrightarrow{c.s.} 0$$

Demostración

Sea $A(\varepsilon) = \{\omega \in \Omega : |X_n(\omega)| > \varepsilon \text{ para una infinidad de valores de } n\}$.

Por la proposición 5.35, $P[A(\varepsilon)] = 0$ para cualquier $\varepsilon > 0$. Así que el resultado se sigue aplicando la proposición 5.33. ■

COROLARIO 5.37. *Sean X, X_1, X_2, \dots variables aleatorias tales que:*

$$\sum_{n=1}^{\infty} P[|X_n - X| > \varepsilon] < \infty$$

para cualquier $\varepsilon > 0$.

Entonces:

$$X_n \xrightarrow{c.s.} X$$

5.4. Funciones generadoras y convergencia en distribución

TEOREMA 5.38. *Sea X_1, X_2, \dots una sucesión de variables aleatorias que admiten como posibles valores únicamente enteros no negativos y sean Φ_1, Φ_2, \dots sus correspondientes funciones generadoras de probabilidades, entonces el límite $\lim_{n \rightarrow \infty} P[X_n = k]$ existe para cualquier $k \in \{0, 1, \dots\}$ si y sólo si el límite $\lim_{n \rightarrow \infty} \Phi_n(t)$ existe para cualquier $t \in (0, 1)$. Además, en ese caso, si $f(k) = \lim_{n \rightarrow \infty} P[X_n = k]$ y $\Phi(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$, entonces $\Phi(t) = \sum_{k=0}^{\infty} f(k)t^k$.*

Demostración

Supongamos primero que $\lim_{n \rightarrow \infty} P[X_n = k]$ existe para cualquier $k \in \{0, 1, \dots\}$. Para $k \in \{0, 1, \dots\}$, $n \in \mathbb{N}$ y $t \in (0, 1)$, definamos $f(k) = \lim_{n \rightarrow \infty} P[X_n = k]$, $f_n(k) = P[X_n = k]$ y $\Phi(t) = \sum_{k=0}^{\infty} f(k)t^k$. Obsérvese que como $0 \leq f(k) \leq 1$ para cualquier k , Φ está bien definida.

Fijemos $t \in (0, 1)$ y sea $r \in \mathbb{N}$. Se tiene entonces:

$$\begin{aligned} |\Phi_n(t) - \Phi(t)| &= \left| \sum_{k=0}^{\infty} [f_n(k) - f(k)] t^k \right| \leq \sum_{k=0}^{\infty} |f_n(k) - f(k)| t^k \\ &= \sum_{k=0}^{r-1} |f_n(k) - f(k)| t^k + \sum_{k=r}^{\infty} |f_n(k) - f(k)| t^k \\ &\leq \sum_{k=0}^{r-1} |f_n(k) - f(k)| t^k + \sum_{k=r}^{\infty} t^k \end{aligned}$$

Como la serie $\sum_{k=1}^{\infty} t^k$ es convergente, dada $\varepsilon > 0$ existe $r \in \mathbb{N}$ tal que $\sum_{k=r}^{\infty} t^k < \frac{\varepsilon}{2}$.

Ahora, como $\lim_{n \rightarrow \infty} f_n(k) = f(k)$ para cualquier $k \in \{0, 1, \dots\}$, entonces existe $N \in \mathbb{N}$ tal que $|f_n(k) - f(k)| < \frac{\varepsilon}{2r}$ para cualquier $k \in \{0, \dots, r-1\}$ y $n \geq N$.

Por lo tanto, para $n \geq N$, se tiene:

$$|\Phi_n(t) - \Phi(t)| \leq \sum_{k=0}^{r-1} |f_n(k) - f(k)| t^k + \sum_{k=r}^{\infty} t^k < \frac{\varepsilon}{2r} \sum_{k=0}^{r-1} |f_n(k) - f(k)| + \frac{\varepsilon}{2} \leq \varepsilon$$

Así que:

$$\lim_{n \rightarrow \infty} \Phi_n(t) = \Phi(t)$$

lo cual demuestra la primera parte.

Supongamos ahora que $\lim_{n \rightarrow \infty} \Phi_n(t)$ existe para cualquier $t \in (0, 1)$. Para $t \in (0, 1)$, $k \in \{0, 1, \dots\}$ y $n \in \mathbb{N}$, definamos $f_n(k) = P[X_n = k]$, $\Phi_n^{(0)} = \Phi_n$ y $\Phi_n^{(k+1)}(t) = \frac{\Phi_n^{(k)}(t) - f_n(k)}{t}$

Vamos a demostrar, por inducción, que, para cualquier $k \in \{0, 1, \dots\}$, se tienen las siguientes dos propiedades:

- (i) $\Phi_n^{(k)}(t) = f_n(k) + \sum_{j=k+1}^{\infty} f_n(j) t^{j-k}$ para cualquier $t \in (0, 1)$ y $n \in \mathbb{N}$.
- (ii) $\lim_{n \rightarrow \infty} \Phi_n^{(k)}(t)$ existe para cualquier $t \in (0, 1)$ y la función $\Phi^{(k)} = \lim_{n \rightarrow \infty} \Phi_n^{(k)}$ es no decreciente en el intervalo $(0, 1)$.
- (iii) $\lim_{n \rightarrow \infty} f_n(k) = \lim_{t \rightarrow 0^+} \Phi^{(k)}(t)$

Para $k = 0$, se tiene:

$$\Phi_n^{(0)}(t) = \Phi_n(t) = \sum_{k=0}^{\infty} f_n(k) t^k = f_n(0) + \sum_{k=1}^{\infty} f_n(k) t^k$$

Además, por hipótesis, $\lim_{n \rightarrow \infty} \Phi_n^{(0)}(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$ existe para cualquier $t \in (0, 1)$ y, como, para cualquier $n \in \mathbb{N}$, Φ_n es una función no decreciente en el intervalo $(0, 1)$, entonces la función $\Phi^{(0)} = \lim_{n \rightarrow \infty} \Phi_n^{(0)}$ también lo es.

Ahora bien, para cualquier $t \in (0, 1)$ y $n \in \mathbb{N}$, se tiene:

$$\Phi_n^{(0)}(t) = \Phi_n(t) = \sum_{j=0}^{\infty} f_n(j) t^j = f_n(0) + \sum_{j=1}^{\infty} f_n(j) t^j$$

Así que:

$$\Phi_n(t) \geq f_n(0) = \Phi_n(t) - \sum_{k=1}^{\infty} f_n(k) t^k \geq \Phi_n(t) - \sum_{k=1}^{\infty} t^k = \Phi_n(t) - \frac{t}{1-t}$$

De manera que, tomando límites cuando n tiende a ∞ , se obtiene:

$$\Phi(t) \geq \limsup_{n \rightsquigarrow \infty} f_n(0) \geq \liminf_{n \rightsquigarrow \infty} f_n(0) \geq \Phi(t) - \frac{t}{1-t}$$

Finalmente, tomando límites cuando t tiende a 0 por la derecha:

$$\lim_{t \rightsquigarrow 0^+} \Phi(t) \geq \limsup_{n \rightsquigarrow \infty} f_n(0) \geq \liminf_{n \rightsquigarrow \infty} f_n(0) \geq \lim_{t \rightsquigarrow 0^+} \Phi(t)$$

Así que, $\lim_{n \rightsquigarrow \infty} f_n(0) = \lim_{t \rightsquigarrow 0^+} \Phi(t)$.

Supongamos ahora que se cumplen las propiedades *i*, *ii* y *iii* para $k = m$, en donde $m \in \{0, 1, \dots\}$, entonces:

$$\begin{aligned} \Phi_n^{(m+1)}(t) &= \frac{\Phi_n^{(m)}(t) - f_n(m)}{t} = \frac{f_n(m) + \sum_{j=m+1}^{\infty} f_n(j)t^{j-m} - f_n(m)}{t} = \frac{\sum_{j=m+1}^{\infty} f_n(j)t^{j-m}}{t} \\ &= f_n(m+1) + \sum_{j=m+2}^{\infty} f_n(j)t^{j-(m+1)} \end{aligned}$$

Como $\lim_{n \rightsquigarrow \infty} \Phi_n^{(m)}(t)$ existe para cualquier $t \in (0, 1)$ y $\lim_{n \rightsquigarrow \infty} f_n(m)$ existe, entonces:

$$\lim_{n \rightsquigarrow \infty} \Phi_n^{(m+1)}(t) = \lim_{n \rightsquigarrow \infty} \frac{\Phi_n^{(m)}(t) - f_n(m)}{t}$$

existe para cualquier $t \in (0, 1)$.

Por la propiedad *i*, la función $\Phi_n^{(m+1)}$ es no decreciente en el intervalo $(0, 1)$, así que la función $\Phi^{(m+1)} = \lim_{n \rightsquigarrow \infty} \Phi_n^{(m+1)}$ también lo es.

Ahora bien, para cualquier $t \in (0, 1)$, $k \in \{0, 1, \dots\}$ y $n \in \mathbb{N}$, se tiene:

$$\Phi_n^{(m+1)}(t) = f_n(m+1) + \sum_{j=m+2}^{\infty} f_n(j)t^{j-(m+1)}$$

Así que:

$$\begin{aligned} \Phi_n^{(m+1)}(t) &\geq f_n(m+1) = \Phi_n^{(m+1)}(t) - \sum_{j=m+2}^{\infty} f_n(j)t^{j-(m+1)} \\ &\geq \Phi_n^{(m+1)}(t) - \sum_{j=m+2}^{\infty} t^{j-(m+1)} = \Phi_n^{(m+1)}(t) - \frac{t}{1-t} \end{aligned}$$

De manera que, tomando límites cuando n tiende a ∞ , se obtiene:

$$\Phi^{(m+1)}(t) \geq \limsup_{n \rightsquigarrow \infty} f_n(m+1) \geq \liminf_{n \rightsquigarrow \infty} f_n(m+1) \geq \Phi^{(m+1)}(t) - \frac{t}{1-t}$$

Finalmente, tomando límites cuando t tiende a 0 por la derecha:

$$\lim_{t \rightsquigarrow 0^+} \Phi^{(m+1)}(t) \geq \limsup_{n \rightsquigarrow \infty} f_n(m+1) \geq \liminf_{n \rightsquigarrow \infty} f_n(m+1) \geq \lim_{t \rightsquigarrow 0^+} \Phi^{(m+1)}(t)$$

Así que, $\lim_{n \rightsquigarrow \infty} f_n(m+1) = \lim_{t \rightsquigarrow 0^+} \Phi^{(m+1)}(t)$.

■

COROLARIO 5.39. Sean X, X_1, X_2, \dots variables aleatorias que admiten como posibles valores únicamente enteros no negativos y sean $\Phi, \Phi_1, \Phi_2, \dots$ sus correspondientes funciones generadoras de probabilidades, entonces $X_n \xrightarrow{D} X$ si y sólo si $\Phi(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$ para cualquier $t \in (0, 1)$.

Demostración

Supongamos que $X_n \xrightarrow{D} X$, entonces $\lim_{n \rightarrow \infty} P[X_n = k] = P[X = k]$ para cualquier $k \in \{0, 1, \dots\}$, así que, por el teorema 5.38, $\lim_{n \rightarrow \infty} \Phi_n(t)$ existe para cualquier $t \in (0, 1)$ y si $\Psi(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$, entonces $\Psi(t) = \sum_{k=0}^{\infty} P[X = k] t^k = \Phi(t)$.

Supongamos ahora que $\Phi(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$ para cualquier $t \in (0, 1)$, entonces, por el teorema 5.38, $\lim_{n \rightarrow \infty} P[X_n = k]$ existe para cualquier $k \in \{0, 1, \dots\}$ y si $f(k) = \lim_{n \rightarrow \infty} P[X_n = k]$, entonces $\sum_{k=0}^{\infty} P[X = k] t^k = \Phi(t) = \sum_{k=0}^{\infty} f(k) t^k$. Así que $\lim_{n \rightarrow \infty} P[X_n = k] = f(k) = P[X = k]$. ■

La demostración del siguiente resultado requiere de resultados no expuestos en este libro, de manera que únicamente se enuncia. Puede consultarse una demostración en Billingsley, P., *Probability and Measure*, John Wiley, 1979.

TEOREMA 5.40. Sean X, X_1, X_2, \dots variables aleatorias y supongamos que sus correspondientes funciones generadoras de momentos, M, M_1, M_2, \dots , están definidas en una vecindad común de 0. Entonces $X_n \xrightarrow{D} X$ si y sólo si $M(t) = \lim_{n \rightarrow \infty} M_n(t)$ en una vecindad de 0.

5.5. Ley débil de los grandes números

La ley débil de los grandes números tiene su origen en el teorema de Bernoulli, publicado en el año 1713, el cual establece que si \mathcal{E} es un experimento aleatorio y A un evento relativo a ese experimento, de probabilidad igual a p , y consideramos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento \mathcal{E} , de tal manera que cada repetición es independiente de las otras, entonces, llamando X_n al número de veces que ocurre el evento A en las primeras n repeticiones del experimento, se tiene $\frac{X_n}{n} \xrightarrow{P} p$.

El teorema de Bernoulli equivale a decir que si X_1, X_2, \dots es una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro p , entonces:

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow{P} p$$

La forma general de este resultado se debe al matemático soviético Pafnuty Lvovich Chebyshev, quien en el año 1867 demostró el siguiente resultado:

PROPOSICIÓN 5.41 (**Chebyshev**). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita. Entonces:*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$$

en donde μ es la esperanza común de X_1, X_2, \dots

Demostración

Para cada $n \in \mathbb{N}$, sea $Y_n = \frac{X_1 + \dots + X_n}{n}$. Entonces Y_n es una variable aleatoria de varianza finita y esperanza μ . De manera que, por la desigualdad de Chebyshev, se tiene:

$$P[|Y_n - \mu| > \varepsilon] \leq \frac{1}{\varepsilon^2} \text{Var}[Y_n] = \frac{\sigma^2}{n\varepsilon^2}$$

en donde σ^2 es la varianza común de X_1, X_2, \dots . Tomando límites cuando $n \rightsquigarrow \infty$ se tiene entonces el resultado. ■

El teorema de Bernoulli admite una generalización en otro sentido: Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, todas con distribución Bernoulli, pero no necesariamente idénticamente distribuidas, entonces $\frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^n p_j \xrightarrow{P} 0$, en donde p_j es el parámetro de X_j . Este resultado se debe a Siméon Denis Poisson, quien lo demostró en el año 1800 y lo bautizó como la ley débil de los grandes números. La forma general de este resultado se debe al matemático soviético Andrei Andreyevich Markov, quien en el año 1880 demostró el siguiente resultado:

PROPOSICIÓN 5.42 (**Markov**). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes de varianza finita tales que $\lim_{n \rightsquigarrow \infty} \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 = 0$. Entonces:*

$$\frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^n \mu_j \xrightarrow{P} 0$$

en donde μ_j es la esperanza X_j .

Demostración

Para cada $n \in \mathbb{N}$, sea $Y_n = \frac{X_1 + \dots + X_n}{n}$. Entonces Y_n es una variable aleatoria de varianza finita y esperanza μ . De manera que, por la desigualdad de Chebyshev, se tiene:

$$P\left[\left|Y_n - \frac{1}{n} \sum_{j=1}^n \mu_j\right| > \varepsilon\right] \leq \frac{1}{n^2 \varepsilon^2} \sum_{j=1}^n \sigma_j^2$$

Tomando límites cuando $n \rightsquigarrow \infty$ se tiene entonces el resultado. ■

EJEMPLO 5.43. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, con funciones de densidad f_1, f_2, \dots , respectivamente, dadas por:

$$f_n(x) = \begin{cases} \frac{1}{2} & \text{si } x \in \left\{n^{\frac{1}{4}}, -n^{\frac{1}{4}}\right\} \\ 0 & \text{en otro caso} \end{cases}$$

Para $n \in \mathbb{N}$, se tiene $\mu_n = 0$ y $\sigma_n^2 = \sqrt{n}$, así que:

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{j=1}^n \sqrt{j} \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sqrt{n} = \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} = 0$$

Por lo tanto, con base en la proposición 5.42, se concluye:

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} 0$$

EJEMPLO 5.44. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, con funciones de densidad f_1, f_2, \dots , respectivamente, dadas por:

$$f_n(x) = \begin{cases} \frac{1}{2} & \text{si } x \in \{n, -n\} \\ 0 & \text{en otro caso} \end{cases}$$

Para $n \in \mathbb{N}$, se tiene $\mu_n = 0$ y $\sigma_n^2 = n^2$, así que:

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{j=1}^n \sigma_j^2 = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{j=1}^n j^2 = \lim_{n \rightarrow \infty} \frac{(n+1)(2n+1)}{6n} = \infty$$

Por lo tanto, no se cumple la condición de Markov, la cual permitiría concluir $\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} 0$.

Obsérvese que se tiene:

$$\frac{1}{n} X_n = \frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^{n-1} X_j = \frac{1}{n} \sum_{j=1}^n X_j - \frac{n-1}{n} \frac{1}{n-1} \sum_{j=1}^{n-1} X_j$$

Así que si se tuviera $\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} 0$, entonces se tendría $\frac{1}{n} X_n \xrightarrow{P} 0$. Pero $|\frac{1}{n} X_n| = 1$ con probabilidad 1.

▲

El resultado de Aleksandr Yakovlevich Khintchine, el cual se demuestra más adelante, muestra que la condición de la proposición 5.42 no es necesaria para la validez de la ley débil.

LEMA 5.45. Si $f : [0, \infty) \mapsto \mathbb{R}$ es una función decreciente y no negativa tal que $\int_0^\infty f(x) dx < \infty$ y (a_n) una sucesión monótona creciente de números reales positivos tal que $\lim_{n \rightarrow \infty} a_n = \infty$, entonces $\lim_{n \rightarrow \infty} a_n f(a_n) = 0$.

Demostración

La sucesión (s_n) , en donde $s_n = \sum_{\{k \in \mathbb{N}: k \leq a_n\}} f(k)$, es monótona no decreciente y se tiene:

$$s_n = \sum_{\{k \in \mathbb{N}: k \leq a_n\}} \int_{k-1}^k f(k) dx \leq \sum_{\{k \in \mathbb{N}: k \leq a_n\}} \int_{k-1}^k f(x) dx \leq \int_0^{a_n} f(x) dx \leq \int_0^{\infty} f(x) dx.$$

Así que (s_n) converge y es, por lo tanto, una sucesión de Cauchy.

Entonces, dada $\varepsilon > 0$ existe un número natural M tal que si $n \geq m \geq M$ entonces $s_n - s_m < \frac{\varepsilon}{2}$, es decir $\sum_{\{k \in \mathbb{N}: a_m < k \leq a_n\}} f(k) < \frac{\varepsilon}{2}$.

Sea ahora N tal que $a_n > 2(a_M + 1)$ para cualquier $n > N$, se tiene entonces, para $n > N$, $a_n - 2(a_M + 1) > 0$ y $(a_n - a_M - 1)f(a_n) \leq \sum_{\{k \in \mathbb{N}: a_M < k \leq a_n\}} f(k) < \frac{\varepsilon}{2}$. Así que:

$$a_n f(a_n) < 2(a_n - a_M - 1)f(a_n) < \varepsilon$$

lo cual prueba el resultado. ■

PROPOSICIÓN 5.46. *Si X es una variable aleatoria de esperanza finita y (a_n) una sucesión monótona creciente de números reales positivos tal que $\lim_{n \rightarrow \infty} a_n = \infty$, entonces:*

$$\lim_{n \rightarrow \infty} a_n P[X > a_n] = \lim_{n \rightarrow \infty} a_n P[X < -a_n] = 0$$

Demostración

Como X tiene esperanza finita, se tiene:

$$\int_0^{\infty} P[X > x] dx = \int_0^{\infty} [1 - F_X(x)] dx < \infty$$

y:

$$\int_0^{\infty} P[X < -x] dx \leq \int_0^{\infty} P[X \leq -x] dx = \int_0^{\infty} F_X(-x) dx < \infty$$

Además, las funciones $x \mapsto P[X > x]$ y $x \mapsto P[X < -x]$ son no negativas y decrecientes en el intervalo $[0, \infty)$.

El resultado se sigue entonces del lema 5.45. ■

LEMA 5.47. *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ y (a_n) una sucesión monótona creciente de números reales positivos tal que $\lim_{n \rightarrow \infty} a_n = \infty$. Para $n, k \in \mathbb{N}$, definamos:*

$$Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}$$

Entonces, fijando n , las variables aleatorias Y_1^n, Y_2^n, \dots tienen la misma distribución. Además, si μ_n es la esperanza común de Y_1^n, Y_2^n, \dots , entonces $\lim_{n \rightarrow \infty} \mu_n = \mu$.

Demostración

$$\begin{aligned} F_{Y_k^n}(x) &= P[Y_k^n \leq x] = P[Y_k^n \leq x, |X_k| \leq a_n] + P[Y_k^n \leq x, |X_k| > a_n] \\ &= P[X_k \leq x, |X_k| \leq a_n] + P[Y_k^n \leq x, |X_k| > a_n] \\ &= \begin{cases} 0 & \text{si } x < -a_n \\ P[-a_n \leq X_k \leq x] & \text{si } -a_n \leq x < 0 \\ P[-a_n \leq X_k \leq x] + P[|X_k| > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases} \\ &= \begin{cases} 0 & \text{si } x < -a_n \\ P[-a_n \leq X_k \leq x] & \text{si } -a_n \leq x < 0 \\ P[X_k \leq x] + P[X_k > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases} \\ &= \begin{cases} 0 & \text{si } x < -a_n \\ F_{X_k}(x) - P[X_k < -a_n] & \text{si } -a_n \leq x < 0 \\ F_{X_k}(x) + P[X_k > a_n] & \text{si } 0 \leq x \leq a_n \\ 1 & \text{si } x > a_n \end{cases} \end{aligned}$$

De manera que, fijando n , las variables aleatorias Y_1^n, Y_2^n, \dots tienen la misma distribución. Además:

$$\begin{aligned} \mu_n &= E[Y_1^n] = \int_0^\infty [1 - F_{Y_1^n}(x)] dx - \int_0^\infty F_{Y_1^n}(-x) dx \\ &= \int_0^{a_n} [1 - F_{Y_1^n}(x)] dx - \int_0^{a_n} F_{Y_1^n}(-x) dx \\ &= \int_0^{a_n} [1 - F_{X_1}(x) - P[X_1 > a_n]] dx - \int_0^{a_n} [F_{X_1}(-x) - P[X_1 < -a_n]] dx \\ &= \int_0^{a_n} [1 - F_{X_1}(x)] dx - a_n P[X_1 > a_n] - \int_0^{a_n} F_{X_1}(-x) dx + a_n P[X_1 < -a_n] \\ &= \int_0^{a_n} [1 - F_{X_1}(x)] dx - \int_0^{a_n} F_{X_1}(-x) dx + a_n P[X_1 < -a_n] - a_n P[X_1 > a_n] \end{aligned}$$

Así que, utilizando la proposición 5.46, $\lim_{n \rightarrow \infty} \mu_n = E[X_1] = \mu$.

■

El siguiente resultado fue demostrado por Aleksandr Yakovlevich Khintchine en el año 1928:

PROPOSICIÓN 5.48 (Khintchine). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ . Entonces:*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$$

Demostración

Sea ν es el valor común de $E[|X_1|], E[|X_2|], \dots$. Si $\nu = 0$, el resultado es trivial. Supongamos entonces que $\nu > 0$.

Dada $\delta > 0$, definamos, para $n, k \in \mathbb{N}$:

$$a_n = \frac{\delta \varepsilon^2}{8\nu} n \text{ y } Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}$$

Por el lema 5.47, fijando n , las variables aleatorias Y_1^n, Y_2^n, \dots tienen la misma distribución y si μ_n es la esperanza común de Y_1^n, Y_2^n, \dots , entonces $\lim_{n \rightarrow \infty} \mu_n = \mu$.

Por otra parte, para cualesquiera $n, k \in \mathbb{N}$, se tiene $(Y_k^n)^2 \leq a_n^2$, así que Y_k^n tiene varianza finita.

Además, $|Y_k^n| \leq |X_k|$ y $|Y_k^n| \leq a_n$, así que, si σ_n^2 es la varianza común de Y_1^n, Y_2^n, \dots , se tiene:

$$\sigma_n^2 \leq E[(Y_k^n)^2] \leq E[a_n |X_k|] = a_n E[|X_k|] = \frac{\delta \varepsilon^2}{8\nu} n E[|X_k|] = \frac{\delta n \varepsilon^2}{8}.$$

Ahora bien, como $\lim_{n \rightarrow \infty} \mu_n = \mu$ y $\lim_{n \rightarrow \infty} a_n P[X_1 > a_n] = 0$, existe N tal que $|\mu_n - \mu| < \frac{\varepsilon}{2}$ y $a_n P[X_1 > a_n] < \frac{\delta^2}{2}$ para cualquier $n > N$.

Entonces, para $n > N$, se tiene:

$$\begin{aligned} & P \left[\left| \frac{X_1 + \dots + X_n}{n} - \mu \right| > \varepsilon \right] \\ & \leq P \left[\left| \frac{Y_1^n + \dots + Y_n^n}{n} - \mu \right| > \varepsilon \right] + P[Y_k^n \neq X_k \text{ para alguna } k \leq n] \\ & \leq P \left[\left| \frac{Y_1^n + \dots + Y_n^n}{n} - \mu_n \right| > \frac{\varepsilon}{2} \right] + P[Y_k^n \neq X_k \text{ para alguna } k \leq n] \end{aligned}$$

Pero, por la desigualdad de Chebyshev, se tiene:

$$P \left[\left| \frac{Y_1^n + \dots + Y_n^n}{n} - \mu_n \right| > \frac{\varepsilon}{2} \right] \leq \frac{4\sigma_n^2}{n\varepsilon^2} \leq \frac{\delta}{2}.$$

Además:

$$\begin{aligned}
P[Y_k^n \neq X_k \text{ para alguna } k \leq n] &\leq \sum_{k=1}^n P[Y_k^n \neq X_k] \\
&= \sum_{k=1}^n P[|X_k| > a_n] = nP[X_1 > a_n] \\
&= \frac{n}{a_n} a_n P[X_1 > a_n] = \frac{1}{\delta} a_n P[X_1 > a_n] < \frac{\delta}{2}
\end{aligned}$$

Así que:

$$P\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \varepsilon\right] \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

lo cual prueba el resultado. ■

El método utilizado por Khintchine en la proposición anterior es conocido como el **método de truncación**. Fue introducido por Markov en el año 1913 con relación a un teorema de Aleksandr Mikhailovich Lyapunov, el cual generaliza el teorema de de Moivre.

EJEMPLO 5.49. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes con función de densidad común f dada por:

$$f(x) = \begin{cases} \frac{c}{x^3} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene $\mu = \sum_{k=1}^{\infty} \frac{c}{k^2} < \infty$, así que: por el teorema de Khintchine, $\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$.

Obsérvese que se tiene $\sigma^2 = \sum_{k=1}^{\infty} \frac{c}{k} - \sum_{k=1}^{\infty} \frac{c}{k^2} = \infty$.

5.5.1. Interpretación de la Esperanza. La ley débil de los grandes números permite recuperar la interpretación de la esperanza de una variable aleatoria como el promedio de los valores que toma ésta cuando el experimento aleatorio se repite muchas veces.

EJEMPLO 5.50. Supongamos que se participa en un juego en el cual la ganancia esperada es de μ pesos, entonces, por la ley débil de los grandes números, dada cualquier $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \leq \varepsilon\right] = 1$$

Esto significa que dada $\delta > 0$, existe N tal que $P\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \leq \varepsilon\right] > 1 - \delta$ para cualquier $n \geq N$, lo cual equivale a decir que, para cualquier $n \geq N$:

$$P[n(\mu - \varepsilon) \leq X_1 + \dots + X_n \leq n(\mu + \varepsilon)] > 1 - \delta$$

Tomemos, por ejemplo, $\mu = 1$, $\varepsilon = 0.01$ y $\delta = 0.01$. Entonces existe N tal que, para cualquier $n \geq N$, se tiene:

$$P[0.99n \leq X_1 + \cdots + X_n \leq 1.01n] > 0.99$$

En particular:

$$P[X_1 + \cdots + X_n \geq 0.99n] > 0.99$$

$$P[X_1 + \cdots + X_n > 1.01n] < 0.01$$

▲

El resultado que obtuvimos en el capítulo 5, del primer volumen de este libro, con relación a la obra Hamlet de Shakespeare, el cual puede parecer sorprendente, puede ahora entenderse un poco mejor. Ahí consideramos la obra Hamlet de Shakespeare, el conjunto C de caracteres tipográficos que ahí se utilizan y el número total T de caracteres físicamente distintos que se utilizan en la obra. Suponiendo entonces que una persona escribe una secuencia de T caracteres, cada uno seleccionado al azar del conjunto C , la probabilidad de que esa secuencia de T caracteres coincida exactamente con la obra de Shakespeare resulta ser igual a $p = \left(\frac{1}{m}\right)^T$, en donde m es el número de elementos que hay en C . Suponiendo ahora que el experimento consistente en escribir una secuencia de T caracteres, cada uno seleccionado al azar del conjunto C , se repite indefinidamente, siendo cada repetición independiente de las demás y definiendo como éxito, en cada repetición de este experimento, al hecho de obtener una secuencia de T caracteres que coincida exactamente con la obra de Shakespeare, tenemos entonces una sucesión de ensayos de Bernoulli, independientes, en cada uno de los cuales la probabilidad de éxito es igual a p . Demostramos entonces que la probabilidad de que en algún momento una de las secuencias de T caracteres coincida exactamente con la obra de Shakespeare es igual a 1. Podemos preguntarnos ahora cuántas secuencias de T caracteres tendrían que escribirse, en promedio, hasta llegar a obtener en algún momento la obra de Shakespeare. La respuesta es la esperanza de la variable aleatoria Y definida como es el número de repeticiones del experimento que se tienen que realizar hasta obtener por primera vez éxito en la mencionada sucesión de ensayos de Bernoulli. Como $Y - 1$ tiene distribución geométrica, se tiene $E[Y] = 1 + \frac{1-p}{p} = \frac{1}{p} = m^T$.

Sin embargo, cabe comentar que el resultado no demuestra la validez de tal interpretación práctica. En primer lugar porque la ley débil establece únicamente que la probabilidad de que el promedio $\frac{X_1 + \cdots + X_n}{n}$ difiera de la esperanza μ en una cantidad grande es pequeña, lo cual no significa que, necesariamente, en una repetición particular del experimento aleatorio correspondiente, el promedio de los valores de

la variable aleatoria se acercará cada vez más a μ . Ni siquiera significa que, realizando muchas secuencias de repeticiones del experimento aleatorio, los casos en que el promedio de los valores de la variable aleatoria no se acerquen cada vez más a μ serán raros, a menos que integremos la interpretación frecuencial de la probabilidad, la cual no queda demostrada tampoco por la ley débil. En otras palabras, la ley débil de los grandes números es un resultado puramente teórico, el cual se obtiene a partir de las propiedades del modelo matemático que hemos considerado hasta este momento. Su interpretación práctica requiere de consideraciones adicionales que no están contenidas dentro del modelo teórico. Sin embargo, hay algo, en conexión con esta discusión, que sí se deriva de la ley débil y es el hecho de que, si bien el resultado no demuestra la validez de la interpretación frecuencial de la probabilidad, sí muestra que el modelo probabilístico que hemos desarrollado es perfectamente compatible con tal interpretación.

Por otra parte, incluso como resultado teórico, debe de tenerse cuidado en la interpretación de lo que dice la ley débil pues se establece únicamente que, cuando n es grande, hay una probabilidad cercana a 1 de que el valor absoluto de la diferencia $\frac{X_1 + \dots + X_n}{n} - \mu$ sea menor que un número positivo fijo de antemano. Esto no debe interpretarse en el sentido de que, cuando n es grande, la suma $X_1 + \dots + X_n$ y la cantidad $n\mu$ difieran en muy poco con una probabilidad muy grande.

Esta idea queda más clara en el contexto de los juegos “justos”. Supongamos para esto que X representa la ganancia que se recibe al participar en un juego. La suma $X_1 + \dots + X_n$ representa entonces la ganancia acumulada en n juegos, mientras que la cantidad $n\mu$ representa el pago total que se debe hacer por participar en los n juegos de tal manera que cada uno de ellos sea justo. ¿Puede esperarse que la ganancia acumulada será aproximadamente igual al pago total? La respuesta es, no necesariamente, lo cual se ilustra en el siguiente ejemplo.

EJEMPLO 5.51. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, todas con función de densidad dada por:*

$$f(x) = \begin{cases} \frac{1}{j(j+1)2^j} & \text{si } x = 2^j \text{ con } j \in \mathbb{N} \\ 1 - \sum_{j=1}^{\infty} \frac{1}{j(j+1)2^j} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Se tiene $E[X_j] = \sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1$. De manera que, por la proposición 5.48, para cualquier $\varepsilon > 0$, se tiene:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{S_n}{n} - 1 \right| > \varepsilon \right] = 0$$

en donde $S_n = \sum_{k=1}^n X_k$.

De aquí se sigue que:

$$\lim_{n \rightarrow \infty} P[|S_n - n| \leq \varepsilon n] = 1$$

lo cual nos dice que cuando n es grande, con probabilidad muy cercana a 1, la ganancia en este juego se encuentra localizada en el intervalo $[-\varepsilon n, \varepsilon n]$. Como puede verse, debido a que dicho intervalo es muy grande, esa relación no nos da una idea clara sobre el valor de S_n .

Se puede dar una mayor precisión en cuanto a la localización de S_n . Para esto, definamos, para cada número natural $n \geq 2$, $a_n = \frac{n}{\log_2 n}$ y, para $k \in \mathbb{N}$:

$$Y_k^n = \begin{cases} X_k & \text{si } |X_k| \leq a_n \\ 0 & \text{en otro caso} \end{cases}$$

Siguiendo la demostración de la proposición 5.48, se tiene que, fijando n , las variables aleatorias Y_1^n, Y_2^n, \dots tienen la misma distribución y su varianza es finita. Además:

$$E[Y_k^n] = \sum_{\{j \in \mathbb{N}: 2^j \leq a_n\}} 2^j P[X_k = 2^j] = \sum_{\{j \in \mathbb{N}: 2^j \leq a_n\}} \frac{1}{j(j+1)}$$

$$\begin{aligned} \text{Var}(Y_k^n) &\leq E[(Y_k^n)^2] = \sum_{\{j \in \mathbb{N}: 2^j \leq a_n\}} 2^{2j} P[X_k = 2^j] \\ &= \sum_{\{j \in \mathbb{N}: 2^j \leq a_n\}} \frac{2^j}{j(j+1)} \leq \sum_{\{j \in \mathbb{N}: 2^j \leq a_n\}} \frac{2^j}{j^2} = \sum_{\{j \in \mathbb{N}: j \leq \log_2 a_n\}} \frac{2^j}{j^2} \end{aligned}$$

Vamos a requerir de una cota superior adecuada para esta última sumatoria. Para obtenerla, sea $z_n = \frac{n^2}{2^n} \sum_{k=1}^n \frac{2^k}{k^2}$, entonces:

$$\begin{aligned} z_{n+1} &= \frac{(n+1)^2}{2^{n+1}} \sum_{k=1}^{n+1} \frac{2^k}{k^2} = \frac{(n+1)^2}{2^{n+1}} \sum_{k=1}^n \frac{2^k}{k^2} + 1 = \frac{(n+1)^2}{2^{n+1}} \frac{2^n}{n^2} z_n + 1 \\ &= \frac{1}{2} \left(1 + \frac{1}{n}\right)^2 z_n + 1 \end{aligned}$$

Supongamos $z_{n+1} \leq z_n$, entonces:

$$z_{n+2} = \frac{1}{2} \left(1 + \frac{1}{n}\right)^2 z_{n+1} + 1 \leq \frac{1}{2} \left(1 + \frac{1}{n}\right)^2 z_n + 1 = z_{n+1}$$

Además, $z_5 = \frac{5^2}{2^5} \sum_{k=1}^5 \frac{2^k}{k^2} = \frac{347}{72}$, $z_4 = \frac{4^2}{2^4} \sum_{k=1}^4 \frac{2^k}{k^2} = \frac{352}{72}$. De manera que $z_5 < z_4$.

Por lo tanto, por el principio de inducción matemática, $z_{n+1} \leq z_n$ para cualquier $n \geq 4$.

Así que la sucesión $(z_n)_{n \geq 4}$ es monótona decreciente y como es no negativa, converge.

Aunque no se requiere tener el valor de $\lim_{n \rightsquigarrow \infty} z_n$, se puede obtener fácilmente. En efecto, sea $z = \lim_{n \rightsquigarrow \infty} z_n$, entonces, como $z_{n+1} = \frac{1}{2} \left(1 + \frac{1}{n}\right)^2 z_n + 1$, tomando límites cuando $n \rightsquigarrow \infty$, se obtiene $z = \frac{1}{2}z + 1$, de lo cual resulta $z = 2$.

Como la sucesión $(z_n)_{n \geq 4}$ converge, está acotada. Sea M una cota superior de $(z_n)_{n \geq 1}$, se tiene entonces:

$$z_n = \frac{n^2}{2^n} \sum_{k=1}^n \frac{2^k}{k^2} \leq M$$

para cualquier $n \in \mathbb{N}$.

Por lo tanto:

$$\sum_{k=1}^n \frac{2^k}{k^2} \leq \frac{M2^n}{n^2} \text{ para cualquier } n \in \mathbb{N}.$$

Sea k_n es el más grande entero k tal que $2^k \leq a_n$, es decir, $2^{k_n} \leq a_n$ y $k_n + 1 > \log_2 a_n$. Entonces, tomando n suficientemente grande de tal manera que $\log_2 a_n \geq 2$, se tiene:

$$\sum_{\{j \in \mathbb{N}: j \leq \log_2 a_n\}} \frac{2^j}{j^2} = \sum_{j=1}^{k_n} \frac{2^j}{j^2} \leq \frac{M2^{k_n}}{k_n^2} \leq \frac{Ma_n}{(\log_2 a_n - 1)^2} \leq \frac{Ma_n}{\left(\frac{1}{2} \log_2 a_n\right)^2} = \frac{4Ma_n}{(\log_2 a_n)^2}.$$

Así que:

$$\text{Var}(Y_k^n) \leq \sum_{\{j \in \mathbb{N}: j \leq \log_2 a_n\}} \frac{2^j}{j^2} \leq \frac{Aa_n}{(\log_2 a_n)^2}$$

en donde $A = 4M$.

$$\text{Sea ahora } b_n = \sum_{k=1}^n E[Y_k^n] = n \sum_{\{k \in \mathbb{N}: 2^k \leq a_n\}} \frac{1}{k(k+1)}.$$

Obsérvese que $\lim_{n \rightsquigarrow \infty} a_n = \infty$, pero como $\lim_{n \rightsquigarrow \infty} \frac{a_n}{n} = \lim_{n \rightsquigarrow \infty} \frac{1}{\log_2 n} = 0$, a_n crece mucho más lento que n . Además, se tiene:

$$\lim_{n \rightsquigarrow \infty} \sum_{\{k \in \mathbb{N}: 2^k \leq a_n\}} \frac{1}{k(k+1)} = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$$

De manera que, cuando n es grande, $b_n \approx n$.

El siguiente resultado localiza entonces de manera más precisa el valor de S_n :

$$\lim_{n \rightsquigarrow \infty} P[|S_n - b_n| \leq \varepsilon a_n] = 1$$

Demostremos la validez de esta última relación. Se tiene:

$$\begin{aligned} & P[|S_n - b_n| > \varepsilon a_n] \\ & \leq P\left[\left|\sum_{k=1}^n E[Y_k^n] - b_n\right| > \varepsilon a_n\right] + P[Y_k^n \neq X_k \text{ para alguna } k \leq n] \end{aligned}$$

Pero, por la desigualdad de Chebyshev y tomando n suficientemente grande de tal manera que $\log_2 a_n \geq 2$, se tiene:

$$\begin{aligned} P[|\sum_{k=1}^n E[Y_k^n] - b_n| > \varepsilon a_n] &\leq \frac{1}{\varepsilon^2 a_n^2} \sum_{k=1}^n \text{Var}(Y_k^n) \\ &= \frac{n}{\varepsilon^2 a_n^2} \text{Var}(Y_1^n) \leq \frac{n}{\varepsilon^2 a_n^2} \frac{A a_n}{(\log_2 a_n)^2} = \frac{A n}{\varepsilon^2 a_n (\log_2 a_n)^2} \end{aligned}$$

Además:

$$\begin{aligned} P[Y_k^n \neq X_k \text{ para alguna } k \leq n] &\leq \sum_{k=1}^n P[Y_k^n \neq X_k] \\ &= \sum_{k=1}^n P[|X_k| > a_n] = n P[|X_1| > a_n] \\ &= n \sum_{\{k \in \mathbb{N}: 2^k > a_n\}} \frac{1}{k(k+1)2^k} \leq n \sum_{\{k \in \mathbb{N}: k > \log_2 a_n\}} \frac{1}{k^2 2^k} \\ &\leq \frac{n}{(\log_2 a_n)^2} \sum_{\{k \in \mathbb{N}: k > \log_2 a_n\}} \frac{1}{2^k} \leq \frac{2n}{a_n (\log_2 a_n)^2} \end{aligned}$$

Así que:

$$P[|S_n - b_n| > \varepsilon a_n] \leq \frac{A n}{\varepsilon^2 a_n (\log_2 a_n)^2} + \frac{2n}{a_n (\log_2 a_n)^2} = \frac{B n}{a_n (\log_2 a_n)^2} = \frac{B}{\log_2 a_n}$$

en donde B es una constante.

Por lo tanto:

$$\lim_{n \rightarrow \infty} P[|S_n - b_n| > \varepsilon a_n] = 0$$

lo cual prueba el resultado.

Ahora bien:

$$b_n = n \sum_{\{k \in \mathbb{N}: 2^k \leq a_n\}} \frac{1}{k(k+1)} = n \left(1 - \frac{1}{k_{k+1}}\right)$$

en donde, como antes, k_n es el más grande entero k tal que $2^k \leq a_n$. Es decir, se tiene $k_n \leq \log_2 a_n$ y $k_n + 1 > \log_2 a_n$.

En particular:

$$k_n + 1 \leq 1 + \log_2 a_n = 1 + \log_2 n - \log_2 \log_2 n < \log_2 n$$

Por otra parte, como $\lim_{n \rightarrow \infty} \frac{\log_2 n}{\log_2 \log_2 n} = 0$, dada $\delta > 0$, existe N tal que si $n > N$, entonces $\log_2 n < \delta \log_2 \log_2 n$.

De esta manera, si $\delta = 1 - \frac{1}{1+\varepsilon}$, existe N tal que si $n > N$, entonces:

$$k_n + 1 > \log_2 a_n = \log_2 n - \log_2 \log_2 n > \log_2 n - \delta \log_2 n = (1 - \delta) \log_2 n = \frac{\log_2 n}{1 + \varepsilon}$$

Por lo tanto:

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[S_n - n \leq -(1 - \varepsilon) \frac{n}{\log_2 n} \right] &= \lim_{n \rightarrow \infty} P \left[S_n - n \leq \varepsilon a_n - \frac{n}{\log_2 n} \right] \\ &\geq \lim_{n \rightarrow \infty} P \left[S_n - n \leq \varepsilon a_n - \frac{n}{k_n + 1} \right] = P \left[S_n - n + \frac{n}{k_n + 1} \leq \varepsilon a_n \right] \\ &\geq P \left[\left| S_n - n + \frac{n}{k_n + 1} \right| \leq \varepsilon a_n \right] = \lim_{n \rightarrow \infty} P \left[|S_n - b_n| \leq \varepsilon a_n \right] = 1. \\ \lim_{n \rightarrow \infty} P \left[S_n - n \geq -(1 + 2\varepsilon) \frac{n}{\log_2 n} \right] &= \lim_{n \rightarrow \infty} P \left[S_n - n \geq -\varepsilon a_n - \frac{n(1 + \varepsilon)}{\log_2 n} \right] \\ &\geq \lim_{n \rightarrow \infty} P \left[S_n - n \leq -\varepsilon a_n - \frac{n}{k_n + 1} \right] = P \left[S_n - n + \frac{n}{k_n + 1} \leq -\varepsilon a_n \right] \\ &\geq P \left[\left| S_n - n + \frac{n}{k_n + 1} \right| \leq \varepsilon a_n \right] = \lim_{n \rightarrow \infty} P \left[|S_n - b_n| \leq \varepsilon a_n \right] = 1 \end{aligned}$$

Así que:

$$\lim_{n \rightarrow \infty} P \left[-(1 + 2\varepsilon) \frac{n}{\log_2 n} \leq S_n - n \leq -(1 - \varepsilon) \frac{n}{\log_2 n} \right] = 1$$

lo cual muestra que, con una probabilidad muy cercana a 1, la ganancia acumulada S_n será considerablemente menor al pago total n , tendiendo a ser la diferencia infinitamente grande.

Obsérvese que este resultado no contradice la ley débil pues, a pesar de que la diferencia $S_n - n$ pueda hacerse infinitamente grande, la diferencia $\frac{S_n}{n} - 1$ se mantiene pequeña. En efecto, se tiene:

$$\lim_{n \rightarrow \infty} P \left[-\frac{(1 + 2\varepsilon)}{\log_2 n} \leq \frac{S_n}{n} - 1 \leq -\frac{(1 - \varepsilon)}{\log_2 n} \right] = 1$$

y la longitud del intervalo $\left[-\frac{(1 + 2\varepsilon)}{\log_2 n}, -\frac{(1 - \varepsilon)}{\log_2 n} \right]$ tiende a 0 cuando n tiende a ∞ .

5.6. Ley fuerte de los grandes números

Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita y esperanza común μ . La ley débil de los grandes números establece que $\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mu$. En el año 1930 Andrey Nikolaevich Kolmogorov mostró que este resultado puede mejorarse demostrando que la convergencia

a μ se da no sólo en probabilidad sino también con probabilidad 1, la cual, como ya vimos, es un tipo de convergencia más fuerte.

Como vimos antes, la demostración de que la sucesión $Y_n = \frac{X_1 + \dots + X_n}{n}$ converge a μ en probabilidad está basada en la desigualdad de Chebyshev, de la cual se obtiene que $P[|Y_n - \mu| > \varepsilon] \leq \frac{K}{n}$, en donde K es una constante. De la proposición 5.36 puede verse, que para demostrar que la sucesión Y_n converge a μ con probabilidad 1 bastaría con demostrar que $\sum_{n=1}^{\infty} P[|Y_n - \mu| > \varepsilon] < \infty$ para cualquier $\varepsilon > 0$. Para probar esto no basta con aplicar la desigualdad de Chebyshev puesta ésta únicamente establece que $P[|Y_n - \mu| > \varepsilon] \leq \frac{K}{n}$ y la serie $\sum_{n=1}^{\infty} \frac{1}{n}$ no es convergente.

El resultado de Kolmogorov tiene su origen en el teorema de Borel, publicado en el año 1909, el cual se enuncia y demuestra a continuación:

PROPOSICIÓN 5.52 (Teorema de Borel). *Sea \mathcal{E} un experimento aleatorio y A un evento relativo a ese experimento, de probabilidad igual a p . Consideremos un nuevo experimento aleatorio consistente en la repetición indefinida del experimento \mathcal{E} , de tal manera que cada repetición es independiente de las otras. Sea X_n el número de veces que ocurre el evento A en las primeras n repeticiones del experimento, entonces:*

$$\frac{X_n}{n} \xrightarrow{\text{c.s.}} p$$

Demostración

Sabemos que X_n tiene distribución binomial de parámetros n y p . Así que:

$$E[X_n] = np$$

$$E[X_n^2] = np + n(n-1)p^2$$

$$E[X_n^3] = np + 3n(n-1)p^2 + n(n-1)(n-2)p^3$$

$$E[X_n^4] = np + 7n(n-1)p^2 + 6n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4$$

Por lo tanto:

$$\begin{aligned} E\left[\left(\frac{X_n}{n} - p\right)^4\right] &= \frac{E[X_n^4]}{n^4} - 4\frac{E[X_n^3]}{n^3}p + 6\frac{E[X_n^2]}{n^2}p^2 - 4\frac{E[X_n]}{n}p^3 + p^4 \\ &= \frac{1}{n^3}p(1-p)[3np(1-p) - 6p(1-p) + 1] < \frac{1}{4n^3}\left(\frac{3n}{4} + n\right) < \frac{1}{n^2} \end{aligned}$$

Sabemos además que si X es cualquier variable aleatoria y ε cualquier número real positivo, entonces $P[|X| \geq \varepsilon] \leq \frac{1}{\varepsilon}E[|X|]$, así que:

$$P\left[\left|\frac{X_n}{n} - p\right| > \varepsilon\right] \leq \frac{E\left[\left(\frac{X_n}{n} - p\right)^4\right]}{\varepsilon^4} < \frac{1}{n^2\varepsilon^4}$$

La serie $\sum_{n=1}^{\infty} P \left[\left| \frac{X_n}{n} - p \right| > \varepsilon \right]$ es entonces convergente para cualquier $\varepsilon > 0$. Así que, por el corolario 5.36, $\frac{X_n}{n} - p \xrightarrow{c.s.} 0$, es decir, $\frac{X_n}{n} \xrightarrow{c.s.} p$.

El teorema de Borel equivale a decir que si X_1, X_2, \dots es una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro p , entonces $\frac{X_1 + \dots + X_n}{n} \xrightarrow{c.s.} p$.

Rajchman mostró, en el año 1932, que la convergencia con probabilidad 1 se puede establecer demostrándola primero para una subsucesión, como se expone a continuación:

PROPOSICIÓN 5.53 (Rajchman). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de varianza finita. Entonces, para cualquier $\varepsilon > 0$, se tiene:*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{c.s.} \mu$$

en donde μ es la esperanza común de X_1, X_2, \dots

Demostración

Para cada $n \in \mathbb{N}$, sea $S_n = \sum_{k=1}^n (X_k - \mu)$ y $Y_n = \frac{S_n}{n}$. Entonces Y_n es una variable aleatoria de varianza finita y esperanza 0. De manera que, por la desigualdad de Chebyshev, se tiene:

$$P[|Y_n| > \varepsilon] \leq \frac{1}{\varepsilon^2} E[Y_n^2] = \frac{\sigma^2}{n\varepsilon^2}$$

en donde σ^2 es la varianza común de X_1, X_2, \dots

$$\text{Así que, } \sum_{n=1}^{\infty} P[|Y_n| > \varepsilon] \leq \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2\varepsilon^2} < \infty.$$

Por lo tanto, por la proposición 5.36:

$$P[\lim_{n \rightarrow \infty} Y_n^2 = 0] = 1$$

Sea ahora $Z_n = \max_{\{k: n^2 \leq k < (n+1)^2\}} |S_k - S_{n^2}|$. Entonces:

$$Z_n^2 = \max_{\{k: n^2 \leq k < (n+1)^2\}} |S_k - S_{n^2}|^2 \leq \sum_{k=n^2}^{(n+1)^2-1} |S_k - S_{n^2}|^2$$

Así que:

$$E[|S_k - S_{n^2}|^2] = E\left[\left|\sum_{j=n^2+1}^k (X_j - \mu)\right|^2\right] = \sum_{j=n^2+1}^k E[(X_j - \mu)^2]$$

$$\leq \sum_{j=n^2+1}^{(n+1)^2-1} E[(X_k - \mu)^2] = 2n\sigma^2$$

Por lo tanto:

$$E[Z_n^2] \leq \sum_{k=n^2}^{(n+1)^2-1} E[|S_k - S_{n^2}|^2] \leq (2n+1)2n\sigma^2 \leq 6n^2\sigma^2$$

De manera que, por la desigualdad de Chebyshev, se tiene:

$$P\left[\left|\frac{Z_n}{n^2}\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} E\left[\frac{Z_n^2}{n^4}\right] \leq \frac{6n^2\sigma^2}{n^4\varepsilon^2} = \frac{6\sigma^2}{n^2\varepsilon^2}$$

$$\text{Así que, } \sum_{n=1}^{\infty} P\left[\left|\frac{Z_n}{n^2}\right| > \varepsilon\right] \leq \sum_{n=1}^{\infty} \frac{6\sigma^2}{n^2\varepsilon^2} < \infty.$$

Por lo tanto, por la proposición 5.36:

$$P\left[\lim_{n \rightarrow \infty} \frac{Z_n}{n^2} = 0\right] = 1$$

$$\text{Sea } A = \left\{\omega \in \dot{\Omega} : \lim_{n \rightarrow \infty} \frac{Z_n(\omega)}{n^2} = 0\right\} \text{ y } B = \left\{\omega \in \dot{\Omega} : \lim_{n \rightarrow \infty} Y_{n^2}(\omega) = 0\right\}.$$

$$\text{Se tiene } P(A \cap B) = 1 - P(A^c \cup B^c) \geq 1 - P(A^c) - P(B^c) = 1.$$

Además, si $\omega \in A \cap B$, dada $\varepsilon > 0$ existe $N(\omega) \in \mathbb{N}$ tal que $\frac{Z_n(\omega)}{n^2} < \frac{\varepsilon}{2}$ y $|Y_{n^2}(\omega)| < \frac{\varepsilon}{2}$ para cualquier $n \geq N(\omega)$.

Ahora bien, para $n \in \mathbb{N}$ y $n^2 \leq k < (n+1)^2$:

$$|Y_k| = \frac{|S_k|}{k} \leq \frac{|S_{n^2}|}{n^2} + \frac{|S_k - S_{n^2}|}{n^2} = |Y_{n^2}| + \frac{|S_k - S_{n^2}|}{n^2} \leq |Y_{n^2}| + \frac{Z_n}{n^2}$$

Así que, si $\omega \in A \cap B$ y $k \geq [N(\omega)]^2$, entonces $n^2 \leq k < (n+1)^2$ para alguna $n \geq N(\omega)$. Por lo tanto:

$$|Y_k(\omega)| \leq |Y_{n^2}(\omega)| + \frac{Z_n(\omega)}{n^2} < \varepsilon$$

Así que $\lim_{k \rightarrow \infty} Y_k(\omega) = 0$.

Por lo tanto:

$$P\left[\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu\right] = P\left[\lim_{k \rightarrow \infty} Y_k = 0\right] \geq P(A \cap B) = 1$$

■

El método de Kolmogorov para probar la convergencia con probabilidad 1 de la sucesión $Y_n = \frac{X_1 + \dots + X_n}{n}$ es distinto y previo al de Rajchman y tiene además la virtud de ser más general. Su demostración está basada en una desigualdad más general que la de Chebyshev y que él mismo demuestra, por lo cual es llamada la desigualdad de

Kolmogorov. Aquí daremos una versión ligeramente modificada de la demostración original.

PROPOSICIÓN 5.54 (Desigualdad de Kolmogorov). Sean X_1, \dots, X_n n variables aleatorias independientes de varianza finita y ε cualquier número real positivo, entonces:

$$P \left[\max_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \text{Var}[S_n]$$

en donde, para $j \in \{1, \dots, n\}$, $S_j = \sum_{i=1}^j X_i$.

Demostración

Supongamos primero que $E[X_k] = 0$ para cualquier $k \in \{1, \dots, n\}$. Entonces también se tiene $E[S_k] = 0$ para cualquier $k \in \{1, \dots, n\}$.

Sea $A = \left\{ \omega \in \Omega : \max_{1 \leq k \leq n} |S_k(\omega)| > \varepsilon \right\}$ y, para $k \in \{1, \dots, n\}$:

$$A_k = \left\{ \omega \in A : \max_{1 \leq j \leq k-1} |S_j(\omega)| \leq \varepsilon, |S_k(\omega)| > \varepsilon \right\}$$

en donde $\max_{1 \leq j \leq 0} |S_j(\omega)| \equiv 0$.

Entonces, los eventos A_1, \dots, A_n son mutuamente excluyentes y $A = \bigcup_{k=1}^n A_k$. Así que:

$$\begin{aligned} E[S_n^2 I_A] &= E[S_n^2 \sum_{k=1}^n I_{A_k}] = \sum_{k=1}^n E[S_n^2 I_{A_k}] = \sum_{k=1}^n E[(S_k + S_n - S_k)^2 I_{A_k}] \\ &= \sum_{k=1}^n E[(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2) I_{A_k}] \\ &= \sum_{k=1}^n E[S_k^2 I_{A_k}] + 2 \sum_{k=1}^n E[S_k(S_n - S_k) I_{A_k}] + \sum_{k=1}^n E[(S_n - S_k)^2 I_{A_k}] \end{aligned}$$

Pero, por la proposición 1.26 y el corolario 2.40, $S_k I_{A_k}$ y $S_n - S_k$ son independientes y tienen esperanza finita, de manera que, por la proposición 2.41, se tiene:

$$E[S_k I_{A_k} (S_n - S_k)] = E[S_k I_{A_k}] E[S_n - S_k] = E[S_k I_{A_k}] E[S_n - S_k] = 0$$

Por lo tanto:

$$\begin{aligned} \text{Var}[S_n] &= E[S_n^2] \geq E[S_n^2 I_A] = \sum_{k=1}^n E[S_k^2 I_{A_k}] + \sum_{k=1}^n E[(S_n - S_k)^2 I_{A_k}] \\ &\geq \sum_{k=1}^n E[S_k^2 I_{A_k}] \geq \sum_{k=1}^n \varepsilon^2 E[I_{A_k}] = \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P(A) \end{aligned}$$

$$= \varepsilon^2 P \left[\max_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon \right]$$

de lo cual se sigue el resultado.

Para el caso general, sea $Y_k = X_k - E[X_k]$ para $k \in \{1, \dots, n\}$. Entonces, las variables aleatorias Y_1, \dots, Y_n son independientes, tienen varianzas finitas, $\sum_{i=1}^j Y_i = \sum_{i=1}^j (X_i - E[X_i])$ y $E[Y_j] = 0$ para cualquier $j \in \{1, \dots, n\}$. De manera que si ε es cualquier número real positivo y $S_j = \sum_{i=1}^j X_i$ para cualquier $j \in \{1, \dots, n\}$, entonces:

$$\begin{aligned} P \left[\max_{1 \leq j \leq n} |S_j - E[S_j]| > \varepsilon \right] &= P \left[\max_{1 \leq j \leq n} \left| \sum_{i=1}^j Y_i \right| > \varepsilon \right] \\ &\leq \frac{1}{\varepsilon^2} \text{Var} \left[\sum_{i=1}^j Y_i \right] = \frac{1}{\varepsilon^2} \text{Var} [S_n] \end{aligned}$$

■

PROPOSICIÓN 5.55 (Kolmogorov). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes, de varianzas finitas, esperanza nula y tales que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, en donde σ_n^2 es la varianzas de X_n . Entonces:*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{c.s.} \mu$$

Demostración

Para cada $n \in \mathbb{N}$ sea $S_n = \sum_{k=1}^n X_k$ y, para cada $\varepsilon > 0$, sea:

$$A_\varepsilon = \left\{ \omega \in \Omega : \left| \frac{S_n(\omega)}{n} \right| > \varepsilon \text{ para una infinidad de valores de } n \right\}$$

Por la proposición 5.33, para probar el resultado basta con demostrar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$. Para esto definamos:

$$B_{n,\varepsilon} = \left\{ \omega \in \Omega : \left| \frac{S_k(\omega)}{k} \right| > \varepsilon \text{ para alguna } k \in \mathbb{N} \text{ tal que } 2^{n-1} < k \leq 2^n \right\}$$

Evidentemente se tiene:

$$A_\varepsilon = \left\{ \omega \in \Omega : \omega \in B_{n,\varepsilon} \text{ para una infinidad de valores de } n \right\}$$

De manera que, por el lema de Borel-Cantelli, para probar que $P(A_\varepsilon) = 0$ para cualquier $\varepsilon > 0$, basta con demostrar que $\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) < \infty$ para cualquier $\varepsilon > 0$. Pero, utilizando la desigualdad de Kolmogorov, se tiene:

$$P(B_{n,\varepsilon}) = P \left[\max_{2^{n-1} < k \leq 2^n} \left| \frac{S_k}{k} \right| > \varepsilon \right] = P \left[\max_{2^{n-1} < k \leq 2^n} |S_k| > k\varepsilon \right]$$

$$\begin{aligned} &\leq P \left[\max_{2^{n-1} < k \leq 2^n} |S_k| > \varepsilon 2^{n-1} \right] \leq P \left[\max_{1 \leq k \leq 2^n} |S_k| > \varepsilon 2^{n-1} \right] \\ &\leq \frac{1}{\varepsilon^2 2^{2n-2}} \text{Var} [S_{2^n}] = \frac{4}{\varepsilon^2 2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2 \end{aligned}$$

Así que:

$$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{2^{2n}} \sum_{k=1}^{2^n} \sigma_k^2 = \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n \in \mathbb{N}: k \leq 2^n\}} \frac{1}{2^{2n}}$$

Sea ahora n_0 el más pequeño número natural tal que $k \leq 2^{n_0}$, entonces:

$$\sum_{\{n \in \mathbb{N}: k \leq 2^n\}} \frac{1}{2^{2n}} = \sum_{n=n_0}^{\infty} \frac{1}{2^{2n}} = \frac{4}{2^{2n_0}} \leq \frac{4}{k^2}$$

Así que:

$$\sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n \in \mathbb{N}: k \leq 2^n\}} \frac{1}{2^{2n}} \leq 4 \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} < \infty$$

Por lo tanto:

$$\sum_{n=1}^{\infty} P(B_{n,\varepsilon}) \leq \frac{4}{\varepsilon^2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{\{n \in \mathbb{N}: k \leq 2^n\}} \frac{1}{2^{2n}} < \infty$$

■

COROLARIO 5.56. *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes, de varianza finita y tales que $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, en donde σ_n^2 es la varianza de X_n . Entonces:*

$$P \left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (X_k - E[X_k]) = 0 \right] = 1$$

EJEMPLO 5.57. *Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, con funciones de densidad f_1, f_2, \dots , respectivamente, dadas por:*

$$f_n(x) = \begin{cases} \frac{1}{2} & \text{si } x \in \left\{ n^{\frac{1}{4}}, -n^{\frac{1}{4}} \right\} \\ 0 & \text{en otro caso} \end{cases}$$

Para $n \in \mathbb{N}$, se tiene $\mu_n = 0$ y $\sigma_n^2 = \sqrt{n}$, así que:

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^{\frac{3}{2}}} < \infty$$

Por lo tanto, con base en la proposición 5.55, se concluye:

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\text{c.s.}} 0$$

▲

Para el caso en que las variables aleatorias X_1, X_2, \dots sean idénticamente distribuidas se cumple la ley fuerte con la única condición de que la esperanza común de X_1, X_2, \dots

sea finita. La demostración de este resultado se debe también a Kolmogorov y el método de demostración es el de truncación, el cual fue utilizado en la demostración de la ley débil. Se requieren además algunos resultados previos, los cuales se exponen a continuación:

LEMA 5.58. *Sea X una variable aleatoria cualquiera, entonces X tiene esperanza finita si y sólo si la serie $\sum_{n=1}^{\infty} P[|X| \geq n]$ converge.*

Demostración

Se tiene:

$$E[|X|] = \int_0^{\infty} [1 - F_{|X|}(x)] dx = \int_0^{\infty} P[|X| > x] dx = \sum_{n=1}^{\infty} \int_{n-1}^n P[|X| > x] dx$$

Pero:

$$\sum_{n=1}^{\infty} \int_{n-1}^n P[|X| > x] dx \geq \sum_{n=1}^{\infty} \int_{n-1}^n P[|X| \geq n] dx = \sum_{n=1}^{\infty} P[|X| \geq n]$$

$$\sum_{n=1}^{\infty} \int_{n-1}^n P[|X| > x] dx \leq \sum_{n=0}^{\infty} \int_n^{n+1} P[|X| \geq n] dx$$

$$= \int_0^1 P[|X| \geq 0] dx + \sum_{n=1}^{\infty} \int_n^{n+1} P[|X| \geq n] dx = 1 + \sum_{n=1}^{\infty} P[|X| \geq n]$$

De manera que:

$$\sum_{n=1}^{\infty} P[|X| \geq n] \leq E[|X|] \leq 1 + \sum_{n=1}^{\infty} P[|X| \geq n]$$

de lo cual se sigue el resultado. ■

LEMA 5.59. *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita μ . Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

- (i) $\lim_{n \rightarrow \infty} E[Y_n] = \mu$.
- (ii) Y_n tiene varianza finita para cualquier $n \in \mathbb{N}$.
- (iii) $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, en donde σ_n^2 es la varianza de Y_n .
- (iv) $P[\{\omega \in \Omega : \text{existe } N(\omega) \text{ tal que } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}] = 1$

Demostración

i. Se tiene:

$$F_{Y_n}(x) = \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ P[|X_n| > n] + P[-n \leq X_n \leq x] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

$$= \begin{cases} 0 & \text{si } x < -n \\ P[-n \leq X_n \leq x] & \text{si } -n \leq x < 0 \\ 1 - P[x < X_n \leq n] & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

Así que:

$$\begin{aligned} E[Y_n] &= \int_0^\infty [1 - F_{Y_n}(x)] dx - \int_0^n F_{Y_n}(-x) dx \\ &= \int_0^n P[x < X_n \leq n] dx - \int_0^n P[-n \leq X_n \leq -x] dx \\ &= \int_0^n P[x < X_1 \leq n] dx - \int_0^n P[-n \leq X_1 \leq -x] dx \\ &= \int_0^n [1 - F_{X_1}(x)] dx - \int_0^n P[X_1 > n] dx - \int_0^n F_{X_1}(-x) dx + \int_0^n P[X_1 < -n] dx \\ &= \int_0^n [1 - F_{X_1}(x)] dx - \int_0^n F_{X_1}(-x) dx - nP[X_1 > n] + nP[X_1 < -n] \end{aligned}$$

Por lo tanto, utilizando la proposición 5.46, $\lim_{n \rightarrow \infty} E[Y_n] = E[X_1] = \mu$.

ii. Para cualquier $n \in \mathbb{N}$, se tiene $|Y_n| \leq n$, así que Y_n tiene varianza finita.

$$\begin{aligned} \text{iii. } \sum_{n=1}^\infty \frac{\sigma_n^2}{n^2} &\leq \sum_{n=1}^\infty \frac{1}{n^2} E[Y_n^2] = \sum_{n=1}^\infty \frac{1}{n^2} E[X_n^2 I_{\{|X_n| \leq n\}}] \\ &= \sum_{n=1}^\infty \frac{1}{n^2} \sum_{j=1}^n E[X_n^2 I_{[j-1 < |X_n| \leq j]}] \\ &= E[X_1^2 I_{[j-1 < |X_1| \leq j]}] + \frac{1}{2^2} (E[X_1^2 I_{[j-1 < |X_1| \leq j]}] + E[X_2^2 I_{[j-1 < |X_2| \leq j]}]) + \dots \\ &= E[X_1^2 I_{[j-1 < |X_1| \leq j]}] (1 + \frac{1}{2^2} + \dots) + E[X_2^2 I_{[j-1 < |X_2| \leq j]}] (\frac{1}{2^2} + \frac{1}{3^2} + \dots) + \dots \\ &= \sum_{j=1}^\infty E[X_j^2 I_{[j-1 < |X_j| \leq j]}] \sum_{n=j}^\infty \frac{1}{n^2} \end{aligned}$$

Pero, para cualquier $j \in \{2, 3, \dots\}$, se tiene:

$$\sum_{n=j}^\infty \frac{1}{n^2} \leq \int_{j-1}^\infty \frac{1}{x^2} = \frac{1}{j-1} \leq \frac{2}{j}$$

$$\text{Además, } \sum_{n=1}^\infty \frac{1}{n^2} = 1 + \sum_{n=2}^\infty \frac{1}{n^2} \leq 2$$

Así que, $\sum_{n=j}^\infty \frac{1}{n^2} \leq \frac{2}{j}$ para cualquier $j \in \mathbb{N}$.

Además, tomando en cuenta que X_1, X_2, \dots tienen la misma distribución:

$$E [X_j^2 I_{|j-1 < X_j| \leq j}] \leq j E [|X_j| I_{|j-1 < X_j| \leq j}] = j E [|X_1| I_{|j-1 < X_1| \leq j}]$$

Por lo tanto:

$$\begin{aligned} \sum_{j=1}^{\infty} E [X_j^2 I_{|j-1 < X_j| \leq j}] \sum_{n=j}^{\infty} \frac{1}{n^2} &\leq \sum_{j=1}^{\infty} j E [|X_1| I_{|j-1 < X_1| \leq j}] \frac{4}{j} \\ &= 4 \sum_{j=1}^{\infty} E [|X_1| I_{|j-1 < X_1| \leq j}] \end{aligned}$$

Sea ahora $Z_n = \sum_{j=1}^n |X_1| I_{|j-1 < X_1| \leq j}$, entonces la sucesión de variables aleatorias Z_1, Z_2, \dots es monótona no decreciente y $\lim_{n \rightarrow \infty} Z_n(\omega) = |X_1(\omega)|$ para cualquier $\omega \in \Omega$, así que por el corolario 9.40 del primer volumen de este libro:

$$\sum_{j=1}^{\infty} E [|X_1| I_{|j-1 < X_1| \leq j}] = \lim_{n \rightarrow \infty} E [Z_n] = E [|X_1|] < \infty$$

de lo cual se sigue $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$.

$$i.v. P [Y_n \neq X_n] = P [|X_n| > n] = P [|X_1| > n]$$

De manera que, utilizando el lema 5.58:

$$\sum_{n=1}^{\infty} P [Y_n \neq X_n] = \sum_{n=1}^{\infty} P [|X_1| > n] \leq \sum_{n=1}^{\infty} P [|X_1| \geq n] < \infty.$$

Así que, por el lema de Borel-Cantelli, si:

$$A = \{\omega \in \Omega : Y_n(\omega) \neq X_n(\omega) \text{ para una infinidad de valores de } n\}$$

entonces $P(A) = 0$.

Sea ahora:

$$B = \{\omega \in \Omega : \text{existe } N(\omega) \text{ tal que } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}$$

Entonces, $B \supset A^c$, así que, $P(B) \geq P(A) = 1$. ■

COROLARIO 5.60. *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas de esperanza finita. Para $n \in \mathbb{N}$, definamos:*

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

$$P \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [X_k(\omega) - Y_k(\omega)] = 0 \right\} = 1 \right]$$

Demostración

Por la parte *iv* del lema 5.59, si:

$$B = \{\omega \in \Omega : \text{existe } N(\omega) \text{ tal que } Y_n(\omega) = X_n(\omega) \text{ para cualquier } n \geq N(\omega)\}$$

entonces $P(B) = 1$.

Pero si $\omega \in B$, entonces existe $N(\omega)$ tal que $X_n(\omega) - Y_n(\omega) = 0$ para cualquier $n \geq N(\omega)$, así que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [X_k(\omega) - Y_k(\omega)] = 0$$

■

LEMA 5.61. *Sea (x_n) una sucesión convergente de números reales y sea $x = \lim_{n \rightarrow \infty} x_n$. Entonces la sucesión $z_n = \frac{1}{n} \sum_{k=1}^n x_k$ es convergente y $\lim_{n \rightarrow \infty} z_n = x$.*

Demostración

Sea $M > 0$ tal que $|x - x_n| \leq M$ para cualquier $n \in \mathbb{N}$.

Dada $\varepsilon > 0$, sea $m \in \mathbb{N}$ tal que $|x - x_n| < \frac{\varepsilon}{2}$ para cualquier $n \geq m$.

Entonces, para $n > \max\{m, \frac{2mM}{\varepsilon}\}$, se tiene:

$$\begin{aligned} |z_n - x| &= \left| \frac{1}{n} \sum_{k=1}^n x_k - x \right| = \left| \frac{1}{n} \sum_{k=1}^n (x_k - x) \right| \leq \frac{1}{n} \sum_{k=1}^n |x_k - x| \\ &= \frac{1}{n} \sum_{k=1}^m |x_k - x| + \frac{1}{n} \sum_{k=m+1}^n |x_k - x| \\ &\leq \frac{mM}{n} + \frac{(n-m)\varepsilon}{2n} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

lo cual significa que $\lim_{n \rightarrow \infty} z_n = x$.

■

PROPOSICIÓN 5.62 (**Kolmogorov**). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas, de esperanza finita μ . Entonces, para cualquier $\varepsilon > 0$, se tiene:*

$$P \left[\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} = \mu \right] = 1$$

Demostración

Para cada $n \in \mathbb{N}$, sea:

$$Y_n = \begin{cases} X_n & \text{si } |X_n| \leq n \\ 0 & \text{en otro caso} \end{cases}$$

Por el lema 5.59:

Las variables aleatorias Y_1, Y_2, \dots tienen esperanza finita.

$$\lim_{n \rightarrow \infty} E[Y_n] = \mu$$

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$$

en donde σ_n^2 es la varianza de Y_n .

De manera que, por el lema 5.61 y el corolario 5.56, se tiene:

$$\lim \frac{1}{n} \sum_{k=1}^n E[Y_k] = \mu$$

$$P \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (Y_k(\omega) - E[Y_k]) = 0 \right\} \right] = 1$$

de lo cual se obtiene:

$$P \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k(\omega) = \mu \right\} \right] = 1$$

Además, por el corolario 5.60:

$$P \left[\left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [X_k(\omega) - Y_k(\omega)] = 0 \right\} \right] = 1$$

de lo cual se obtiene el resultado. ■

5.7. Teorema de Poisson

El siguiente teorema generaliza el teorema de Poisson, el cual establece que si, para cada $n \in \mathbb{R}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p \in (0, 1)$ de tal manera que $\lambda = np$ es constante, entonces, para cualquier $k \in \{0, 1, \dots\}$, se tiene:

$$\lim_{n \rightarrow \infty} P[X_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

TEOREMA 5.63. *Para cada $n \in \mathbb{N}$, consideremos n ensayos de Bernoulli independientes, X_{n1}, \dots, X_{nn} , tales que la probabilidad de éxito en el k -ésimo ensayo es p_{nk} y supongamos*

- (i) $\lim_{n \rightarrow \infty} \sum_{k=1}^n p_{nk} = \lambda > 0$
- (ii) $\lim_{n \rightarrow \infty} \sum_{k=1}^n p_{nk}^2 = 0$

Definamos $Z_n = \sum_{k=1}^n X_{nk}$. Entonces, para cualquier $k \in \{0, 1, \dots\}$, se tiene:

$$\lim_{n \rightarrow \infty} P[Z_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Demostración

Para cualquier $r \in \mathbb{R}$, se tiene:

$$\Phi_{Z_n}(t) = [1 - p_{n1}(1-t)][1 - p_{n2}(1-t)] \cdots [1 - p_{nn}(1-t)]$$

Así que:

$$\ln \Phi_{Z_n}(t) = \ln [1 - p_{n1}(1-t)] + \ln [1 - p_{n2}(1-t)] + \cdots + \ln [1 - p_{nn}(1-t)]$$

Pero, $\ln(1-x) = -x + o(x)$, así que, dada $\varepsilon > 0$, existe $\delta > 0$ tal que si $0 < x < \delta$ entonces $|o(x)| < \varepsilon x$.

Además, dada $\delta > 0$, existe N tal que si $n \geq N$ entonces $\sum_{k=1}^n p_{nk}^2 < \delta^2$ y, por lo tanto, $p_{nk} < \delta$ para $k \in \{1, \dots, n\}$.

Por lo tanto, si $n \geq N$ y $t \in (0, 1)$, entonces:

$$\ln [1 - p_{nk}(1-t)] = -p_{nk}(1-t) + e_{nk}$$

$$\text{en donde } |e_{nk}| < \varepsilon(1-t)p_{nk} < \varepsilon p_{nk}.$$

Se tiene entonces:

$$\ln \Phi_{Z_n}(t) = -(1-t) \sum_{k=1}^n p_{nk} + \sum_{k=1}^n e_{nk}$$

y:

$$|\sum_{k=1}^n e_{nk}| \leq \sum_{k=1}^n |e_{nk}| < \varepsilon \sum_{k=1}^n p_{nk}$$

Así que, para cualquier $\varepsilon > 0$, se tiene:

$$\limsup_{n \rightarrow \infty} |\sum_{k=1}^n e_{nk}| \leq \varepsilon \lim_{n \rightarrow \infty} \sum_{k=1}^n p_{nk} = \varepsilon \lambda$$

Por lo tanto:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n e_{nk} = 0$$

Y entonces:

$$\lim_{n \rightarrow \infty} \ln \Phi_{Z_n}(t) = -(1-t) \lim_{n \rightarrow \infty} \sum_{k=1}^n p_{nk} = -(1-t)\lambda$$

de lo cual se concluye:

$$\lim_{n \rightarrow \infty} \Phi_{Z_n}(t) = e^{-\lambda(1-t)}$$

■

La condición $\lim_{n \rightarrow \infty} \sum_{k=1}^n p_{nk}^2 = 0$ equivale a $\lim_{n \rightarrow \infty} \max \{p_{nk} : k \in \{1, \dots, n\}\} = 0$, así que el resultado obtenido puede interpretarse diciendo que una variable aleatoria cuyo valor sea igual al número de veces que ocurre un cierto evento en un número grande de experimentos independientes de manera que la probabilidad de tal evento sea uniformemente pequeña, se distribuye aproximadamente como una variable aleatoria tipo Poisson. Variables aleatorias de este tipo son por ejemplo las que nos dan el número de accidentes que tienen los individuos de una población (se supone aquí que la probabilidad de que un individuo tenga un accidente es pequeña, pudiendo ser diferentes estas probabilidades para diferentes individuos), o también las que nos dan el número de llamadas telefónicas que llegan a una oficina en una jornada de trabajo (se supone aquí, por ejemplo, que la probabilidad de que llegue una llamada en cada lapso de tres minutos es pequeña), o también la que nos da el número de personas que solicitan un servicio en una jornada de trabajo suponiendo que la probabilidad de que una persona solicite el servicio en un determinado lapso de tiempo es pequeña.

5.8. Teorema del límite central

El teorema del límite central tiene su origen en el teorema de de Moivre, publicado en el año 1733, el cual establece que si $x \in \mathbb{R}$ y, para cada $n \in \mathbb{R}$, X_n es una variable aleatoria con distribución binomial de parámetros n y $p \in (0, 1)$, entonces:

$$\lim_{n \rightarrow \infty} P \left[\frac{X_n - np}{\sqrt{npq}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}x^2} dx$$

El teorema de de Moivre equivale a decir que si X_1, X_2, \dots es una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro p , entonces:

$$\lim_{n \rightarrow \infty} P \left[\frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}x^2} dx$$

La forma general de este resultado se debe a los trabajos de la llamada escuela rusa, en particular a Chebyshev, Markov y sobre todo a Lyapunov, quien en el año 1900 demostró que si X_1, X_2, \dots es una sucesión de variables aleatorias independientes e idénticamente distribuidas, con tercer momento finito, entonces:

$$\lim_{n \rightarrow \infty} P \left[a < \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

en donde μ y σ^2 son la esperanza y varianza común, respectivamente, de X_1, X_2, \dots

Más tarde, en 1922, Lindeberg demostró que si X_1, X_2, \dots es una sucesión de variables aleatorias independientes de varianza finita (no necesariamente idénticamente distribuidas) y tales que:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^2} E \left[(X_k - \mu_k)^2 I_{|X_k - \mu_k| > \varepsilon s_n} \right] = 0$$

para cualquier $\varepsilon > 0$, en donde $s_n^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$ y μ_k y σ_k^2 son la esperanza y varianza, respectivamente, de X_k , entonces:

$$\lim_{n \rightarrow \infty} P \left[a < \frac{X_1 + \dots + X_n - (\mu_1 + \dots + \mu_n)}{\sqrt{\sigma_1^2 + \dots + \sigma_n^2}} < b \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

Obsérvese que la condición de Lindeberg se cumple en particular cuando las variables aleatorias X_1, X_2, \dots son idénticamente distribuidas. En efecto, en ese caso, si X es una variable aleatoria con la misma distribución común de X_1, X_2, \dots y μ y σ^2 son la esperanza y varianza, respectivamente, de X , entonces:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^2} E \left[(X_k - \mu_k)^2 I_{|X_k - \mu_k| > \varepsilon s_n} \right] \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n\sigma^2} E \left[(X - \mu)^2 I_{|X - \mu| > \varepsilon \sigma \sqrt{n}} \right] \\ &= \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} E \left[(X - \mu)^2 I_{|X - \mu| > \varepsilon \sigma \sqrt{n}} \right] \\ &= \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} E \left[(X - \mu)^2 - (X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}} \right] \\ &= \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} \left(\sigma^2 - E \left[(X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}} \right] \right) \\ &= 1 - \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} E \left[(X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}} \right] \end{aligned}$$

Pero la sucesión de variables aleatorias (no negativas) $Y_n = (X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}}$ es monótona creciente y su límite es $(X - \mu)^2$, así que:

$$\lim_{n \rightarrow \infty} E \left[(X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}} \right] = E \left[(X - \mu)^2 \right] = \sigma^2$$

Por lo tanto:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{s_n^2} E \left[(X_k - \mu_k)^2 I_{|X_k - \mu_k| > \varepsilon s_n} \right] \\ &= 1 - \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} E \left[(X - \mu)^2 I_{|X - \mu| \leq \varepsilon \sigma \sqrt{n}} \right] = 0 \end{aligned}$$

A continuación damos una demostración directa de este corolario del resultado de Lindeberg para el caso en que la función generadora de momentos de X existe en una vecindad de 0.

PROPOSICIÓN 5.64 (Teorema del límite central). *Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas de varianza finita. Entonces:*

$$\lim_{n \rightarrow \infty} P \left[\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

en donde μ y σ^2 son la esperanza y varianza común, respectivamente, de X_1, X_2, \dots

Demostración

Asumiremos que la función generadora de momentos de X_i existe en una vecindad de 0. Sea $Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ y φ la función generadora de momentos común de X_1, X_2, \dots , entonces:

$$M_{Z_n}(t) = E[e^{tZ_n}] = E[e^{\frac{t}{\sigma\sqrt{n}}(X_1 + X_2 + \dots + X_n - n\mu)}] = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n \exp\left\{-\frac{n\mu t}{\sigma\sqrt{n}}\right\}$$

Así que:

$$\ln M_{Z_n}(t) = n \ln \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) - \frac{n\mu t}{\sigma\sqrt{n}} = n \left[\ln \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) - \frac{\mu t}{\sigma\sqrt{n}} \right]$$

Por lo tanto, utilizando la regla de l'Hôpital, se tiene:

$$\begin{aligned} \lim_{n \rightarrow \infty} \ln M_{Z_n}(t) &= \frac{t}{2\sigma} \lim_{n \rightarrow \infty} \sqrt{n} \left[\frac{1}{\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)} \varphi'\left(\frac{t}{\sigma\sqrt{n}}\right) - \mu \right] \\ &= 2 \left(\frac{t}{2\sigma}\right)^2 \lim_{n \rightarrow \infty} \left\{ \frac{1}{\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)} \varphi''\left(\frac{t}{\sigma\sqrt{n}}\right) - \frac{1}{\varphi^2\left(\frac{t}{\sigma\sqrt{n}}\right)} \left[\varphi'\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^2 \right\} \\ &= 2 \left(\frac{t}{2\sigma}\right)^2 \{ \varphi''(0) - [\varphi'(0)]^2 \} = 2 \left(\frac{t}{2\sigma}\right)^2 \sigma^2 = \frac{1}{2}t^2 \end{aligned}$$

De lo cual se concluye:

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{\frac{1}{2}t^2}$$

Así que, por el teorema 5.40, Z_n converge en distribución a una variable aleatoria con función generadora de momentos dada por $M(t) = e^{\frac{1}{2}t^2}$, es decir, a una variable aleatoria X con distribución normal estándar. Esto significa que:

$$\lim_{n \rightarrow \infty} P \left[\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy \quad \blacksquare$$

EJEMPLO 5.65. Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes e idénticamente distribuidas de varianza finita y definamos $S_n = \sum_{k=1}^n X_k$, entonces:

$$P \left[|S_n - \mu_{S_n}| \leq \sigma_{S_n} \right] \approx \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-\frac{1}{2}x^2} dx = 0.6827$$

$$P \left[|S_n - \mu_{S_n}| \leq 2\sigma_{S_n} \right] \approx \frac{1}{\sqrt{2\pi}} \int_{-2}^2 e^{-\frac{1}{2}x^2} dx = 0.9545$$

$$P \left[|S_n - \mu_{S_n}| \leq 3\sigma_{S_n} \right] \approx \frac{1}{\sqrt{2\pi}} \int_{-3}^3 e^{-\frac{1}{2}x^2} dx = 0.9973$$

Así que, la desviación estándar de S_n nos da una idea de que tanto se separa S_n de su valor esperado.

EJEMPLO 5.66. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, todas con distribución exponencial de parámetro $\lambda = 1$. Encuentre el más pequeño valor de n tal que $P \left[\left| \frac{1}{n} \sum_{k=1}^n X_k - 1 \right| \leq 0.01 \right] \geq 0.9$.

Utilizando el teorema del límite central, se tiene:

$$\begin{aligned} P \left[\left| \frac{1}{n} \sum_{k=1}^n X_k - 1 \right| \leq 0.01 \right] &= P \left[\left| \sum_{k=1}^n X_k - n \right| \leq 0.01n \right] \\ &= P \left[\left| \frac{X_1 + \dots + X_n - n}{\sqrt{n}} \right| \leq 0.01\sqrt{n} \right] \approx \frac{1}{\sqrt{2\pi}} \int_{-0.01\sqrt{n}}^{0.01\sqrt{n}} e^{-\frac{1}{2}y^2} dy \geq 0.9 \end{aligned}$$

Así que, $0.01\sqrt{n} = 1.645$, es decir, $n \geq (1.645)^2(10,000) = 27,060$.

EJEMPLO 5.67. 50 números seleccionados aleatoriamente se redondean al entero más cercano y después se suman. Suponiendo que los errores de redondeo están uniformemente distribuidos en el intervalo $(-0.5, 0.5)$, encuentre la probabilidad de que la suma que se obtiene difiera del valor exacto en más de 3 unidades.

Solución

Sean a_1, \dots, a_{50} los 50 números que se redondean y sean X_1, \dots, X_{50} los respectivos errores de redondeo. Se tiene entonces:

$$\begin{aligned} E[X_i] &= 0, \quad Var[X_i] = \frac{1}{12} \\ P \left[\left| \sum_{i=1}^{50} (a_i + X_i) - \sum_{i=1}^{50} a_i \right| > 3 \right] &= P \left[\left| \sum_{i=1}^{50} X_i \right| > 3 \right] \\ &= P \left[\left| \frac{\sum_{i=1}^{50} X_i}{\sqrt{\frac{50}{12}}} \right| > \frac{3}{\sqrt{\frac{50}{12}}} \right] = P \left[\left| \frac{\sum_{i=1}^{50} X_i}{\sqrt{\frac{50}{12}}} \right| > 1.4697 \right] \\ &\approx \frac{2}{\sqrt{2\pi}} \int_{1.4697}^{\infty} e^{-\frac{1}{2}y^2} dy = 0.14164 \end{aligned}$$

EJEMPLO 5.68. Aplique el teorema del límite central a una sucesión de variables aleatorias independientes, todas con distribución Poisson con el mismo parámetro, para demostrar que:

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}$$

Solución

Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, todas con distribución Poisson de parámetro $\lambda = 1$. Entonces $S_n = X_1 + \dots + X_n$ tiene distribución Poisson de parámetro n , así que:

$$P[S_n \leq n] = e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$$

Por otra parte, por el teorema del límite central, se tiene:

$$\begin{aligned} \lim_{n \rightarrow \infty} P[S_n \leq n] &= \lim_{n \rightarrow \infty} P\left[\frac{S_n - n}{\sqrt{n}} \leq 0\right] = \lim_{n \rightarrow \infty} P\left[\frac{S_n - \mu_{S_n}}{\sigma_{S_n}} \leq 0\right] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 e^{-\frac{1}{2}x^2} dx = \frac{1}{2} \end{aligned}$$

Así que:

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = \frac{1}{2}$$

5.9. Convergencia de series aleatorias

TEOREMA 5.69. Sea X_1, X_2, \dots una sucesión de variables aleatorias, independientes, de varianza finita y tales que las series $\sum_{n=1}^{\infty} \mu_n$ y $\sum_{n=1}^{\infty} \sigma_n^2$ convergen, en donde μ_n y σ_n^2 son la esperanza y la varianza de X_n respectivamente. Entonces, con probabilidad 1, la serie $\sum_{n=1}^{\infty} X_n$ converge.

Demostración

Sea $Z_n = \sum_{j=1}^n X_j - \sum_{j=1}^n \mu_j$, entonces, por la desigualdad de Kolmogorov, para cualquier $\varepsilon > 0$ y $k, r \in \mathbb{N}$, se tiene:

$$\begin{aligned} &P\left[\max_{k < n \leq k+r} |Z_n - Z_k| > \varepsilon\right] \\ &= P\left[\max_{k < n \leq k+r} \left|\sum_{j=k+1}^n X_j - \sum_{j=k+1}^n \mu_j\right| > \varepsilon\right] \leq \frac{1}{\varepsilon^2} \sum_{j=k+1}^{k+r} \sigma_j^2 \leq \frac{1}{\varepsilon^2} \sum_{j=k+1}^{\infty} \sigma_j^2 \end{aligned}$$

Así que:

$$P \left[\sup_{n>k} |Z_n - Z_k| > \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{j=k+1}^{\infty} \sigma_j^2$$

Por lo tanto:

$$\begin{aligned} & \lim_{k \rightsquigarrow \infty} P \left[\bigcap_{n=k+1}^{\infty} [|Z_n - Z_k| \leq \varepsilon] \right] \\ &= \lim_{k \rightsquigarrow \infty} P \left[\sup_{n>k} |Z_n - Z_k| \leq \varepsilon \right] = 1 - \lim_{k \rightsquigarrow \infty} P \left[\sup_{n>k} |Z_n - Z_k| > \varepsilon \right] \\ &\geq 1 - \lim_{k \rightsquigarrow \infty} \frac{1}{\varepsilon^2} \sum_{j=k+1}^{\infty} \sigma_j^2 = 1 \end{aligned}$$

Para cada $r \in \mathbb{N}$, sea $B_{k,r} = \bigcap_{n=k}^{\infty} [|Z_n - Z_k| \leq \frac{1}{r}]$, entonces, por lo demostrado arriba, se tiene $\lim_{k \rightsquigarrow \infty} P[B_{k,r}] = 1$.

Sean ahora $B_r = \bigcup_{k=1}^{\infty} B_{k,r}$ y $B = \bigcap_{r=1}^{\infty} B_r$.

Si $\omega \in B$, entonces $\omega \in B_r$ para cualquier $r \in \mathbb{N}$, de manera que, para cualquier $r \in \mathbb{N}$, existe $k \in \mathbb{N}$ tal que $|Z_n(\omega) - Z_k(\omega)| \leq \frac{1}{r}$ para cualquier $n \geq k$.

Por otra parte, dada $\varepsilon > 0$ existe $r \in \mathbb{N}$ tal que $\frac{1}{r} < \frac{\varepsilon}{2}$, así que, si $\omega \in A$, existe $k \in \mathbb{N}$ tal que $|Z_n(\omega) - Z_k(\omega)| \leq \frac{1}{r} < \frac{\varepsilon}{2}$ para cualquier $n \geq k$. De manera que, para cualesquiera $n, m \geq k$, se tiene:

$$|Z_n(\omega) - Z_m(\omega)| \leq |Z_n(\omega) - Z_k(\omega)| + |Z_m(\omega) - Z_k(\omega)| < \varepsilon$$

Por lo tanto, si $\omega \in B$, la sucesión $(Z_n(\omega))$ es de Cauchy, así que converge.

Ahora bien, fijando r , la sucesión $B_{k,r}$ es monótona creciente, así que:

$$P(B_r) = \lim_{k \rightsquigarrow \infty} P(B_{k,r}) = 1$$

Además, la sucesión B_r es monótona decreciente, así que:

$$P(B) = \lim_{r \rightsquigarrow \infty} P(B_r) = 1$$

■

EJEMPLO 5.70. Sea (c_n) una sucesión de números reales y (X_n) una sucesión de variables aleatorias independientes, todas con función de densidad f dada por:

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } x \in \{1, -1\} \\ 0 & \text{en otro caso} \end{cases}$$

Consideremos la serie $\sum_{n=1}^{\infty} c_n X_n$.

Definiendo $Y_n = c_n X_n$, se tiene $\mu_{Y_n} = 0$ y $\sigma_{Y_n}^2 = c_n^2$, así que si la serie $\sum_{n=1}^{\infty} c_n^2$ es convergente, entonces, con probabilidad 1, la serie $\sum_{n=1}^{\infty} c_n X_n$ converge.

EJEMPLO 5.71. Sea (c_n) una sucesión de números reales positivos y (X_n) una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro $p = \frac{1}{2}$.

Consideremos la serie $\sum_{n=1}^{\infty} c_n X_n$.

Definiendo $Y_n = c_n X_n$, se tiene $\mu_{Y_n} = \frac{1}{2}c_n$ y $\sigma_{Y_n}^2 = \frac{1}{4}c_n^2$, así que si la serie $\sum_{n=1}^{\infty} c_n$ converge, entonces, con probabilidad 1, la serie $\sum_{n=1}^{\infty} c_n X_n$ converge. Este resultado no dice nada que no se sepa de antemano pues si la serie $\sum_{n=1}^{\infty} c_n$ converge, entonces eliminando cualquier colección de c_n 's, la serie que se forma sigue siendo convergente.

Supongamos ahora que la serie $\sum_{n=1}^{\infty} c_n$ no es convergente y consideremos la serie $\sum_{n=1}^{\infty} (c_n X_n - \frac{1}{2}c_n)$. Definiendo $Y_n = c_n X_n - \frac{1}{2}c_n$, se tiene $\mu_{Y_n} = 0$ y $\sigma_{Y_n}^2 = \frac{1}{4}c_n^2$, así que si la serie $\sum_{n=1}^{\infty} c_n^2$ converge, entonces, con probabilidad 1, la serie $\sum_{n=1}^{\infty} (c_n X_n - \frac{1}{2}c_n)$ converge.

Este resultado implica, en particular, que la serie $\sum_{n=1}^{\infty} c_n X_n$ no es convergente, pues si lo fuera, la serie $\sum_{n=1}^{\infty} c_n$ también sería convergente.

Se concluye entonces que, si la serie $\sum_{n=1}^{\infty} c_n^2$ es convergente, la serie $\sum_{n=1}^{\infty} c_n X_n$ es convergente con probabilidad 1 si y sólo si la serie $\sum_{n=1}^{\infty} c_n$ converge.

EJERCICIOS

EJERCICIO 5.1. Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{P} X$ y $\lim_{n \rightarrow \infty} E[(X_n - Y_n)^2] = 0$. Demuestre que $Y_n \xrightarrow{P} X$.

EJERCICIO 5.2. Sea $\Omega = (0, 1]$ y P la medida de Lebesgue. Definamos $X : \Omega \rightarrow \mathbb{R}$ por $X(\omega) = \omega$ y, para cada $n \in \mathbb{N}$, $X_n = \sum_{k=1}^n \frac{k}{n} I_{(\frac{k-1}{n}, \frac{k}{n}]}$. Demuestre directamente que X_n converge a X en probabilidad y casi seguramente.

EJERCICIO 5.3. Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{D} X$ y $Y_n \xrightarrow{D} 0$. Demuestre que $X_n + Y_n \xrightarrow{D} X$.

EJERCICIO 5.4. Sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y, para cada $n \in \mathbb{N}$, X_n una variable aleatoria con distribución uniforme en el conjunto $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$. a) Demuestre directamente que X_n converge a X en distribución. b) Demuestre que la sucesión de funciones generadoras M_{X_n} converge a M_X .

EJERCICIO 5.5. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes tales que $E[X_i] = 0$ para toda i y $\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = 0$, en donde σ_i^2 denota la varianza de X_i . Demuestre que:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{X_1 + \dots + X_n}{n} \right| > \varepsilon \right] = 0$$

para cualquier $\varepsilon > 0$.

EJERCICIO 5.6. Sea Y_1, Y_2, \dots una sucesión de variables aleatorias independientes, todas con distribución Bernoulli. Utilice el resultado del ejercicio 5.5 para demostrar que:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{Y_1 + \dots + Y_n}{n} - \frac{p_1 + \dots + p_n}{n} \right| \leq \varepsilon \right] = 1$$

para cualquier $\varepsilon > 0$, en donde p_i denota el parámetro de la distribución de Y_i .

EJERCICIO 5.7. Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, con funciones de densidad f_1, f_2, \dots , respectivamente, dadas por:

$$f_n(x) = \begin{cases} \frac{1}{2^{n+1}} & \text{si } x \in \{2^n, -2^n\} \\ 1 - \frac{1}{2^n} & \text{si } x = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Demuestre que la sucesión X_1, X_2, \dots no satisface la condición de Markov, pero $\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{P} 0$.

Sugerencia: Utilice el método de truncación.

EJERCICIO 5.8. Suponga que el cambio diario en el precio de una acción de una compañía, en el mercado de valores, es una variable aleatoria con esperanza 0 y varianza σ^2 . Es decir, si, para cualquier $n \in \mathbb{N}$, Y_n es el precio de la acción en el día n , entonces $Y_n = Y_{n-1} + X_n$, en donde X_1, X_2, \dots son variables aleatorias independientes, idénticamente distribuidas, con esperanza 0 y varianza σ^2 . Consideremos una acción cuyo precio el día de hoy es de 100 y para la cual $\sigma^2 = 1$, ¿qué se puede decir acerca de la probabilidad de que en cada uno de los siguientes 10 días el precio de la acción permanecerá entre 95 y 105?

EJERCICIO 5.9. Para cada $n \in \mathbb{N}$, consideremos n variables aleatorias independientes, X_{n_1}, \dots, X_{n_n} , todas con distribución geométrica de parámetros p_{n_1}, \dots, p_{n_n} , respectivamente, y supongamos:

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1-p_{nk}}{p_{nk}} = \lambda > 0$$

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \left(\frac{1-p_{nk}}{p_{nk}} \right)^2 = 0$$

Definiendo $Z_n = \sum_{k=1}^n X_{nk}$, demuestre que, para cualquier $k \in \{0, 1, \dots\}$, se tiene:

$$\lim_{n \rightarrow \infty} P[Z_n = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

EJERCICIO 5.10. Consideremos un experimento aleatorio consistente en lanzar un dado 100 veces y definamos X_i como el resultado del i -ésimo lanzamiento. Encuentre una estimación de $P[\prod_{i=1}^{100} X_i \leq a^{100}]$.

EJERCICIO 5.11. Sean X_1, \dots, X_{20} veinte variables aleatorias independientes, todas con distribución Poisson de parámetro $\lambda = 1$.

a) Obtenga una cota superior para $P[\sum_{i=1}^{20} X_i > 15]$.

b) Utilice el teorema del límite central para estimar $P[\sum_{i=1}^{20} X_i > 15]$.

EJERCICIO 5.12. Sean X_1, \dots, X_{100} cien variables aleatorias independientes, todas con distribución uniforme en el intervalo $(-1, 1)$. Estime la probabilidad:

$$P[30 \leq X_1^2 + \dots + X_{100}^2 \leq 35]$$

EJERCICIO 5.13. Sean X_1, \dots, X_n n variables aleatorias independientes, todas con distribución exponencial de parámetro λ y sea Φ la función de distribución de una variable aleatoria con distribución normal estándar. Utilice el teorema del límite central para expresar la función de distribución de $Z = X_1^2 + \dots + X_n^2$ en términos de Φ .

EJERCICIO 5.14. Suponga que el peso W (en unidades de 1,000 kilos) que un cierto puente puede soportar, sin sufrir daños estructurales, es una variable aleatoria con distribución normal de esperanza 200 y desviación estándar 20. Suponga además que el peso X (en unidades de 1,000 kilos) de un automóvil es una variable aleatoria con esperanza 1 y desviación estándar 0.1. ¿Cuántos automóviles tendrían que estar sobre el puente para que la probabilidad de que sufra daños estructurales exceda 0.1?

EJERCICIO 5.15. Se tienen 100 componentes, los cuales se utilizan en secuencia, es decir: primero se utiliza el componente 1; al fallar éste se utiliza el componente 2; al fallar este último se utiliza el componente 3 y así sucesivamente. Estime la probabilidad de que el tiempo total de vida de los 100 componentes exceda 1,200 suponiendo que, para cada $i \in \{1, \dots, 100\}$, el tiempo de vida del componente i tiene distribución
a) uniforme en el intervalo $(0, 20 + \frac{i}{5})$, b) exponencial con parámetro $\lambda_i = \frac{1}{10 + \frac{i}{10}}$.

EJERCICIO 5.16. Se tienen 60 componentes, los cuales se utilizan en secuencia, es decir, primero se utiliza el componente 1; al fallar éste se utiliza el componente 2; al fallar este último se utiliza el componente 3 y así sucesivamente. Supongamos que, para cada $i \in \{1, \dots, 60\}$, el tiempo de vida del componente i tiene distribución uniforme en el intervalo $(0, 20 + i)$. Estime la probabilidad de que el tiempo total de vida de los 60 componentes no exceda 1,400.

EJERCICIO 5.17. Sea X una variable aleatoria con distribución gama de parámetros $\alpha = 50$ y $\lambda = 5$. Utilice el teorema del límite central para estimar $P[9 \leq X \leq 12]$.

EJERCICIO 5.18. Sea X una variable aleatoria con distribución gama de parámetros $\alpha = 50$ y $\lambda = 5$. Utilice el teorema del límite central y la desigualdad de Chebyshev para estimar $P[|X - 10| \leq 2]$.

EJERCICIO 5.19. Una emisora de radio funciona con una batería, la cual tiene un tiempo de vida distribuido exponencialmente y con esperanza de 1 mes. Cuando una batería se acaba, inmediatamente es sustituida por otra de las mismas características. Encuentre el número mínimo de baterías que se requieren para que, con probabilidad mayor o igual a 0.99, la emisora funcione ininterrumpidamente por lo menos durante un año.

EJERCICIO 5.20. Aplique el teorema del límite central a una sucesión de variables aleatorias independientes, todas con distribución Poisson con el mismo parámetro, para demostrar que

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^{2n} \frac{n^k}{k!} = 1$$

EJERCICIO 5.21. Sea (c_n) una sucesión de números reales y (X_n) una sucesión de variables aleatorias independientes, todas con función de densidad f dada por:

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

en donde $0 < p < 1$.

Demuestre que si la serie $\sum_{n=1}^{\infty} c_n^2$ es convergente, entonces la serie $\sum_{n=1}^{\infty} c_n X_n$ converge con probabilidad 1.

EJERCICIO 5.22. Sea (c_n) una sucesión de números reales positivos tal que la serie $\sum_{n=1}^{\infty} c_n^2$ es convergente y (X_n) una sucesión de variables aleatorias independientes, todas con distribución Bernoulli de parámetro p . Demuestre que la serie $\sum_{n=1}^{\infty} c_n X_n$ es convergente con probabilidad 1 si y sólo si la serie $\sum_{n=1}^{\infty} c_n$ converge.

EJERCICIO 5.23. Sea ξ_1, ξ_2, \dots una sucesión de variables aleatorias independientes cada una de las cuales puede tomar únicamente los valores 0 y 1, cada uno de ellos con probabilidad $\frac{1}{2}$. Demuestre que la variable aleatoria $X = \sum_{k=1}^{\infty} \frac{\xi_k}{2^k}$ tiene distribución uniforme en el intervalo $(0, 1)$.

Parte 3

HISTORIA

CAPÍTULO 6

SURGIMIENTO DEL CÁLCULO DE PROBABILIDADES

Le 29 août 1654

Monsieur,

Nos coups fourrés continuent toujours, et je suis aussi bien que vous dans l'admiration de quoi nos pensées s'ajustent si exactement, qu'il me semble qu'elles aient pris une même route et fait un même chemin: vos derniers traités du Triangle Arithmétique et de son application, en sont une preuve authentique; et si mon calcul ne me trompe, votre onzième conséquence courrait la poste de Paris à Toulouse, pendant que ma proposition des nombres figurés, qui en effet est la même, allait de Toulouse à Paris. Je n'ai garde de faillir, tandis que je rencontrerai de cette sorte; et je suis persuadé que le vraie moyen pour s'empêcher de faillir est celui de concourir avec vous. Mais si j'en disais davantage, la chose tiendrait du compliment, et nous avons banni cet ennemi des conversations douces et aisées.

Carta de Fermat a Pascal

El surgimiento del Cálculo de Probabilidades, como disciplina matemática independiente, tiene como base las soluciones que, durante el periodo que va del año 1654 al año 1657, dieron Blaise Pascal, Pierre de Fermat y Christiaan Huygens a varios problemas, entre los cuales destacan los siguientes:

PROBLEMA 1 (Problema de la división de apuestas). *¿Cómo deben repartirse las apuestas en un juego que se interrumpe? Por ejemplo, suponiendo que dos jugadores, A y B, apuestan 32 pesos cada uno en un juego que consiste de partidas consecutivas, en cada una de las cuales cada jugador tiene la misma posibilidad de ganarla y quien la gane acumula un punto, de tal manera que el juego es ganado por quien obtenga*

primero cuatro puntos, ¿cómo deben de repartirse las apuestas en caso de que el juego se interrumpa cuando el jugador A ha ganado dos puntos y B un punto?

PROBLEMA 2. ¿Cuántas veces se necesita lanzar un dado para que sea más favorable obtener por lo menos un 6?

PROBLEMA 3. ¿Cuántas veces se necesita lanzar un par de dados para que sea más favorable obtener por lo menos un par de seises?

PROBLEMA 4. ¿Cuántos dados se requieren lanzar para que sea más favorable obtener por lo menos dos seises?

PROBLEMA 5. Dos jugadores, P y Q , juegan a lanzar alternadamente un par de dados. El juego comienza lanzando P el par de dados, con la condición de que si obtiene una suma igual a 6 gana el juego, en caso contrario el juego continua lanzando Q el par de dados, con la condición de que si obtiene una suma igual a 7 gana el juego, en caso contrario el juego continua lanzando P el par de dados bajo las condiciones iniciales. ¿Cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?

PROBLEMA 6 (**Problema de la ruina del jugador**). Dos jugadores, A y B , los cuales poseen 12 fichas cada uno, juegan a lanzar sucesivamente tres dados, estableciéndose que A dará una ficha a B cada vez que se obtenga una suma igual a 11, mientras que B dará una ficha a A cada vez que se obtenga una suma igual a 14. Si el ganador del juego es el primero que llegue a poseer las 24 fichas, ¿cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?

Sin embargo, debe señalarse que no son Pascal, Fermat y Huygens los primeros en resolver problemas de probabilidad. De hecho, el análisis de sus soluciones a los problemas planteados muestra que la idea de calcular una probabilidad como un cociente entre casos favorables y casos totales era ya conocida en el medio científico. Previamente a su trabajo se tenía ya estudiado el número de maneras en que puede resultar el lanzamiento de dos o tres dados y se habían resuelto algunos problemas simples relacionados con este resultado. También era ya aceptado en la época de Pascal, Fermat y Huygens que existe una relación entre el número de casos que favorecen la ocurrencia de un evento y la frecuencia con que éste se observa. De hecho, previo al trabajo de Pascal, Fermat y Huygens, existía ya un estudio sistemático sobre el Cálculo de Probabilidades, el cual fue realizado por Girolamo Cardano en el año 1526 en un libro titulado “Liber de Ludo Aleae”, cuya primer publicación apareció en el año 1663.

Uno de los objetivos en este capítulo será el tratar de ubicar correctamente la contribución de Pascal-Fermat-Huygens en la Teoría de la Probabilidad, pues si bien es cierto que no son ellos los primeros en plantear y resolver correctamente problemas de probabilidad, si es su trabajo el que mayor influencia tuvo en el desarrollo posterior del Cálculo de Probabilidades.

6.1. Algunos resultados particulares

Una de las características de un experimento aleatorio es su posibilidad de diferentes resultados. En este sentido, aquellos problemas en donde se trate de determinar las diferentes posibilidades de ocurrencia de un experimento aleatorio pueden considerarse ya como problemas de probabilidad.

Los primeros problemas de este tipo que se plantearon se refieren a lanzamientos de dados. Dado un cierto número de dados, se trataba de encontrar las diferentes formas en que pueden caer. El planteamiento de este problema se remonta al siglo X, sin embargo, el primer cálculo correcto conocido se ubica en el siglo XIII, este cálculo se refiere al caso de 3 dados y se encuentra contenido en un poema titulado “De Vetula” y escrito por Richard de Fournival (1200-1250). Ahí se afirma que 3 dados pueden caer en un total de 216 caminos. Resulta interesante observar que este número no se obtiene ahí como el producto 6^3 , sino considerando primero los 56 posibles casos no ordenados que se obtienen de la suma $6+30+20$, cuyos términos corresponden al caso de 3 números iguales, de dos iguales y uno distinto y de 3 distintos respectivamente; finalmente se obtiene $216 = (6)(1) + (30)(3) + (20)(6)$.

La primera referencia conocida a una relación entre las diferentes posibilidades de ocurrencia de un evento y la frecuencia con que éste se observa se encuentra en los comentarios a una publicación de “La Divina Comedia” que en el año 1477 hace Benvenuto d’Imola. Dice ahí:

“Concerniente a estos lanzamientos (de dados) debe observarse que los dados son cuadrados y cualquier cara puede caer, así que un número que pueda aparecer en más caminos debe ocurrir más frecuentemente, como en el siguiente ejemplo: con tres dados, tres es el más pequeño número que puede obtenerse y sólo se obtiene con tres ases; cuatro puede obtenerse sólo en un camino, con un dos y dos ases”.

Como puede verse, en la cita hay 3 elementos, primero se hace referencia a la simetría de los dados, lo cual justifica la equiprobabilidad de cada cara; en seguida se da la relación entre el número de formas en que una cierta suma puede obtenerse y la frecuencia con que ésta se observa; finalmente se encuentra el número de caminos en que se puede obtener la suma 3 y la suma 4. D’Imola considera, erróneamente, sólo los casos no ordenados cuando que si cada cara de un dado representa un resultado equiprobable, la frecuencia de ocurrencia de determinado evento depende del total de casos ordenados que lo producen.

La misma referencia a la relación que hay entre el número de caminos en que puede obtenerse una cierta suma al lanzar 3 dados y la frecuencia con que esta suma se observa se encuentra también en un trabajo escrito por Galileo Galilei alrededor del

año 1620, siendo su publicación hasta el año 1718 ([4]). En ese trabajo Galileo se propuso esclarecer una confusión que existía al establecer la relación entre el número de caminos en que se obtiene una cierta suma al lanzar 3 dados y la frecuencia con que ésta se obtiene. Se preguntaba concretamente por qué si 9, 10, 11 y 12 pueden obtenerse en igual número de caminos¹, los jugadores de dados, con base en numerosas observaciones, consideraban 10 y 11 más ventajosos que 9 y 12. Obsérvese que al comparar el número de caminos en que se obtiene cada suma, consideraba sólo los casos no ordenados. La confusión la aclaraba Galileo haciendo ver que, de los caminos señalados, aquellos en los que hay 3 números iguales se obtienen de una sola manera, aquellos en los que hay 2 números iguales y uno distinto se obtienen de 3 maneras y aquellos en los que hay 3 números distintos se obtienen de 6 maneras; así que, en realidad, 10 y 11 pueden obtenerse de 27 maneras distintas, mientras que 9 y 12 solo pueden hacerlo de 25 maneras distintas. En otras palabras, Galileo hacía ver que la relación entre la frecuencia de las diferentes sumas debe establecerse en base a los casos ordenados y no en base a los no ordenados como se pensaba.

En otro trabajo, Galileo hizo un estudio cualitativo de los errores que se cometen en las mediciones astronómicas, considerando que estos errores son inherentes al proceso de medición, es decir, considerando al proceso de medición como un fenómeno aleatorio. Este trabajo adquirió más tarde, ya en la época de aplicaciones del Cálculo de Probabilidades, una gran importancia.

Terminaremos esta parte con la formulación de un problema, el cual adquiriría una gran importancia en el trabajo de Pascal-Fermat-Huygens; nos referimos al llamado problema de la división de apuestas, el cual se encontraba planteado muchos años antes del trabajo de Pascal-Fermat-Huygens. En el libro titulado “Summa de Arithmetica, Geometria, Proportioni et Proportionalità”, escrito por Luca Paccioli en 1487 y publicado en 1494, este problema se encuentra formulado como sigue:

Dos personas juegan de manera que se requiere un total de 60 puntos para ganar, siendo el premio de 22 ducados. Por alguna circunstancia, cuando uno tiene 50 puntos y el otro 30, no pueden continuar el juego. ¿Qué parte del premio le corresponde a cada uno?

Paccioli consideraba que la parte que corresponde a cada uno debe ser proporcional a los puntos que lleva ganados; en este caso, la repartición debe hacerse en la proporción de 5 : 3, es decir, al que lleva 50 puntos le corresponden $\frac{5}{8}(22)$ y al otro $\frac{3}{8}(22)$.

¹ 9 : (6, 2, 1), (5, 3, 1), (5, 2, 2), (4, 4, 1), (4, 3, 2), (3, 3, 3)
 10 : (6, 3, 1), (6, 2, 2), (5, 4, 1), (5, 3, 2), (4, 4, 2), (4, 3, 3)
 11 : (6, 4, 1), (6, 3, 2), (5, 3, 3), (5, 4, 2), (5, 5, 1), (4, 4, 3)
 12 : (6, 5, 1), (6, 4, 2), (6, 3, 3), (5, 5, 2), (5, 4, 3), (4, 4, 4)

Paccioli consideró también el mismo problema para el caso de 3 jugadores, siendo análoga la solución que daba.

Obsérvese que la solución que dio Paccioli deja ver que no estaba considerando lo azaroso del juego pues para la repartición sólo considera los puntos ganados por cada uno, cuando que realmente lo que debe contar para la repartición son los puntos que le faltan a cada uno y las posibilidades de obtenerlos antes que el otro.

6.2. El Trabajo de Girolamo Cardano

El primer estudio sistemático de problemas de probabilidad se debe a Girolamo Cardano ([2]). En su trabajo, Cardano realizó un estudio de problemas relacionados con lanzamientos de dados.

En su libro, Cardano trató el problema de determinar el número de posibilidades en el lanzamiento de 2 y 3 dados, obteniendo 36 y 216 respectivamente. Dio además las siguientes tablas en las cuales se expresa el número de caminos en que una cierta suma puede obtenerse con 2 y 3 dados, respectivamente.

Caso de dos dados:

2	12	3	11	4	10	5	9	6	8	7
1		2		3		4		5		6

Caso de tres dados:

3	18	4	17	5	16	6	15	7	14
1		3		6		10		15	
8	13	9	12	10	11				
21		25		27					

Así, por ejemplo, de la primera tabla puede verse que, con dos dados, una suma 6 o una suma 8 puede obtenerse en 5 caminos, de la segunda tabla puede verse que, con 3 dados, una suma 9 o una suma 12 puede obtenerse en 25 caminos.

Aunque en un lenguaje distinto al que se usó más tarde en el Cálculo de Probabilidades, Cardano planteó y resolvió, a la manera clásica, problemas de probabilidad. Un ejemplo es el siguiente:

Considerando el lanzamiento de dos dados, estableció que por lo menos un as se obtiene de 11 maneras; lo mismo puede decirse de por lo menos un dos y así sucesivamente. Agregaba que, sin embargo, un as o un dos no se obtiene de 22 maneras, pues hay 11 maneras en que se obtiene por lo menos un as y 9 más en que se obtiene por lo menos un dos, así que en total son 20 maneras de obtener por lo menos un as o por lo menos un dos. Continuaba diciendo que si se agrega ahora el 3, habrá 7 maneras más y así sucesivamente, en el siguiente paso habrá que sumar 5 maneras más, luego 3 y por último 1.

Decía entonces que si alguien dijera, quiero un as un dos o un tres, se sabe que hay 27 caminos favorables y como el circuito (i.e. todas las posibilidades) es de 36, los caminos en que no se obtiene ninguno de estos números son 9; las posibilidades son entonces de 3 a 1.

Con este razonamiento Cardano llegó de hecho a la llamada definición clásica de probabilidad estableciendo las posibilidades de obtener un determinado resultado en función del número de posibles maneras en que ese resultado puede obtenerse..

Más aún, a partir del resultado anterior, Cardano estableció lo que es un juego justo dándole una interpretación frecuencial a la proporción de las posibilidades. Dice que en cuatro lanzamientos de un par de dados, si la fortuna fuera igual, un as, un dos o un tres caerán 3 veces y sólo una vez no caerá ninguno de ellos, entonces si el jugador que quiere un as, un dos o un tres ganara 3 ducados y el otro 1, en los 4 lanzamientos ganan lo mismo, pues el primero gana una vez y el segundo tres veces.

Estableció entonces una regla general para determinar la apuesta que debe hacer cada jugador de manera que se juegue en igualdad de circunstancias. Según esta regla, las apuestas deben estar en la misma proporción que las posibilidades que cada uno tiene de ganar.

Aparentemente Cardano no dio la correcta interpretación frecuencial de las posibilidades pues no considera en lo anterior un número grande de lanzamientos; sin embargo, el tomar 4 lanzamientos parece ser solo una simplificación para facilitar el razonamiento, pues ya antes, en otra parte de su trabajo, dice que los cálculos (de las diferentes posibilidades) son conjeturas que dan solo una aproximación, pero que en el caso de muchos circuitos (i.e. en el caso de muchos lanzamientos) lo que sucede es muy cercano a la conjetura.

Nótese que, según esto último, Cardano no solo estableció que mientras más posibilidades haya de obtener cierta suma más frecuente será ésta, sino que además afirma que la conjetura es muy cercana a la frecuencia cuando el número de lanzamientos es muy grande.

De lo anterior vemos entonces que Cardano parece ser el primero en dar las bases para una formulación de la definición clásica de probabilidad y el primero en dar una interpretación frecuencial completa a la proporción de posibilidades de ocurrencia de un evento y en definir a partir de esta interpretación lo que se entiende por un juego justo. Todo esto, claro está, no para un experimento aleatorio general, sino para el caso particular de lanzamiento de dados. Pero, como veremos en lo que resta de este capítulo, parece ser que el trabajo de Cardano no tuvo ninguna influencia en el desarrollo del Cálculo de Probabilidades.

El trabajo de Cardano contiene otras consideraciones, algunas de ellas erróneas, aunque encerrando también ideas importantes. Decía, por ejemplo, en una serie de 3 lanzamientos de un dado, hay 91 caminos favorables para al menos un as; entonces, si alguien quiere un as en cada una de dos series de 3 lanzamientos de un dado, multiplicando 91×91 y 125×125 se obtiene 8281 y 15625, respectivamente, así que las posibilidades son aproximadamente de 2 a 1.

El razonamiento de Cardano en este problema parece correcto, tiene dos experimentos independientes cada uno de los cuales consiste en el lanzamiento de 3 dados, entonces los casos favorables a la ocurrencia de un as en cada una de las dos series se obtiene multiplicando 91×91 ; el mismo razonamiento nos da los casos desfavorables, solo que Cardano comete un error en este cálculo pues él toma como casos desfavorables solo aquellos en que no obtiene as en la primera serie y tampoco obtiene as en la segunda serie, cuando que, además de esos, también son desfavorables aquellos en que no se obtiene as en la primera y sí en la segunda y aquellos en que se obtiene un as en la primera y no se obtiene as en la segunda; resultando en total $125 \times 125 + 2 \times 91 \times 125 = 38375$ casos desfavorables; es decir que las posibilidades de obtener un as en cada una de dos series de 3 lanzamientos de un dado son de 8281 a 38375, es decir, aproximadamente de 2 a 9.

En otro libro, titulado “*Practica Arithmeticae Generalis*”, publicado en 1539, Cardano consideró el problema de la división de apuestas. Ahí hizo ver el error de Paccioli al no tomar en cuenta los juegos que faltan por ganar a cada jugador.

6.3. El trabajo de Pascal-Fermat-Huygens

Pascal y Fermat entraron en la escena del Cálculo de Probabilidades cuando en el año 1654 Antoine Gombaud de Méré, conocido como el chevalier de Méré, planteó a Pascal dos problemas; uno relativo a lanzamientos de dados y el problema de la división de apuestas. Pascal resolvió esos problemas y, siendo muy amigo de Fermat, le envió los problemas con sus soluciones, sin darle a conocer sus métodos. Fermat respondió a Pascal con su propia solución a cada uno de los dos problemas planteados, coincidiendo

éstas con las soluciones que había encontrado Pascal. Con esto se iniciaría una serie de reflexiones de Pascal y Fermat acerca de estos problemas, las cuales están contenidas en la correspondencia que sostuvieron durante el año 1654 ([3]).

Desafortunadamente, no toda esa correspondencia logró conocerse; en particular, hasta la fecha se desconoce el método que cada uno siguió para resolver el problema relativo a lanzamientos de dados. Excepto por un artículo escrito por Pascal, titulado “Traité du triangle Arithmétique” ([6]), en el cual desarrolla su método de solución al problema de la división de apuestas, Pascal y Fermat no dieron a conocer sus métodos de solución a los diferentes problemas que se plantearon. Sin embargo, era sabido que ambos habían resuelto un nuevo tipo de problemas y se conocían también los problemas con sus soluciones.

Christiaan Huygens, de origen holandés, entró en escena cuando en el año 1655 visitó Francia y se enteró de los problemas que Pascal y Fermat habían resuelto. Huygens se abocó entonces a la solución de éstos desarrollando un método propio, pues, como ya dijimos, los métodos de Pascal y Fermat eran desconocidos. Huygens publicó el resultado de sus investigaciones en el año 1657 en un libro titulado “De ratiociniis in Ludo Aleae” ([59]). Este trabajo, ignorado por un tiempo, tendría después una gran influencia en el desarrollo del Cálculo de Probabilidades.

En esta sección vamos a analizar tanto el trabajo de Pascal- Fermat como el de Huygens, basándonos para el primero en la correspondencia que se conserva y en el tratado sobre el Triángulo Aritmético de Pascal; para el segundo en el libro de Huygens.

Para el análisis de estos trabajos no seguiremos estrictamente el orden cronológico pues parece más conveniente tratar por separado cada uno de los problemas.

6.3.1. Problema de la división de apuestas. El problema de la división de apuestas o **problema de las partidas**, como lo llamaba Pascal, es el problema 1. Como ya hemos visto, el planteamiento de este problema era ya conocido desde antes, incluso era ya identificado como un problema en donde interviene al azar; pero, hasta antes de Pascal-Fermat, nadie había encontrado una solución correcta a este problema.

Comenzaremos analizando el método usado por Fermat para resolver este problema. Éste se puede descomponer en tres pasos:

1o. Determinar el número máximo de partidas que deberían jugarse a partir de la situación dada para que el juego se termine.

Por ejemplo, en el problema planteado, al jugador P le faltan dos partidas para ganar y al jugador Q le faltan tres partidas. Entonces, a lo más en 4 partidas adicionales se acaba el juego pues no se acaba en las 3 siguientes cuando de éstas P gana una y Q gana dos; pero estando en esta situación, el juego se decide en la siguiente partida.

2o. Suponiendo que se juega el número de partidas encontrado en el primer paso, determinar todos los posibles resultados.

En el ejemplo considerado, se requiere determinar todos los posibles resultados de 4 partidas. Si denotamos por la letra a el que P gane una partida y por la letra b el que gane Q, los posibles resultados en 4 partidas son los siguientes:

$(a, a, a, a), (a, a, a, b), (a, a, b, a), (a, b, a, a), (b, a, a, a)$

$(a, a, b, b), (a, b, a, b), (a, b, b, a), (b, a, a, b), (b, a, b, a), (b, b, a, a)$

$(a, b, b, b), (b, a, b, b), (b, b, a, b), (b, b, b, a), (b, b, b, b)$

en donde, por ejemplo, (b, b, a, b) significa que P gana solo la tercera partida y Q las otras 3. Resultan en total $2^4 = 16$ posibles resultados.

3o. De todos los posibles resultados encontrados en el segundo paso, determinar el número de aquellos que hacen ganar al primer jugador y el número de aquellos que hacen ganar al segundo. La proporción de las apuestas que corresponde a cada uno debe ser entonces igual a la proporción de estos números.

En el ejemplo considerado, hay 11 posibles resultados que hacen ganar al jugador P, a saber, $(a, a, a, a), (a, a, a, b), (a, a, b, a), (a, b, a, a), (b, a, a, a), (a, a, b, b), (a, b, a, b), (a, b, b, a), (b, a, a, b), (b, a, b, a), (b, b, a, a)$. Los 5 restantes hacen ganar al jugador Q. Por lo tanto, las apuestas se las deben repartir en la proporción 11 : 5.

El método de Fermat recibió dos objeciones. La primera de ellas fue hecha por Roberval, amigo de Pascal y según parece el único que conoció los métodos usados por Pascal y Fermat. Con relación al ejemplo que hemos considerado, decía Roberval que no es necesario que se jueguen las 4 partidas pues el juego puede terminarse antes, así que la hipótesis de que se jueguen las 4 partidas es ficticia y debía demostrarse que no es falsa. A esta objeción de Roberval respondió Pascal diciendo que la hipótesis es solo una convención que no afecta el resultado pues si alguno gana el juego en menos de 4 partidas los resultados de las partidas siguientes no afectan porque el otro jugador no podrá llegar a completar los puntos que le faltan para ganar.

La segunda objeción la hizo Pascal a raíz de que Fermat afirmaba que su método era válido no sólo para el caso de dos jugadores sino para cualquier número de ellos. Pascal negaba esta afirmación pues en el caso de 3 jugadores, por ejemplo, la hipótesis

ficticia de extender el juego hasta cierto número de partidas causa problemas: Si por ejemplo al primer jugador le falta una partida, al segundo dos y al tercero también dos, entonces el juego se acaba en a lo más 3 partidas, pero algunos de los posibles resultados de 3 partidas son favorables a dos de los jugadores, por ejemplo, con la misma notación usada para el caso de dos jugadores, el resultado (a, b, b) es favorable tanto al primer jugador como al segundo.

La respuesta que dio Fermat a las dos objeciones planteadas en de singular importancia, en ella Fermat hizo explícita por primera vez la condición necesaria para poder aplicar la definición clásica de probabilidad, es decir, la hipótesis de equiprobabilidad². Además, en la misma respuesta, Fermat introdujo implícitamente la regla de la suma para el caso de eventos mutuamente excluyentes.

Decía Fermat que la ficción de extender el juego hasta un cierto número de partidas no sirve más que para facilitar la regla y **para hacer todos los azares iguales**. Por ejemplo, en el caso planteado por Pascal de los 3 jugadores, en donde al primero le falta una partida para ganar, al segundo dos y al tercero dos, decía Fermat que de los $3^3 = 27$ posibles resultados de 3 partidas, las que son favorables a uno de los jugadores son solo aquellas que lo hacen ganar antes que al otro; de esta manera un posible resultado como (a, b, b) ya no es ambiguo. Haciendo esta consideración encontró que la repartición de las apuestas debe ser en la proporción 17 : 5 : 5. Y, para que no hubiera objeción, dio la siguiente solución, en la cual ya no hay nada ficticio y muestra que, efectivamente, la hipótesis de extender el juego sirve para hacer todos los azares iguales (i.e. equiprobables).

El primer jugador puede ganar en una partida o en dos o en tres. Para ganar en una partida tiene $\frac{1}{3}$ de los azares, para ganar en dos partidas (exactamente) puede hacerlo de dos maneras, así que tiene $\frac{2}{9}$ de los azares, para ganar en tres partidas (exactamente) puede hacerlo de dos maneras, así que tiene $\frac{2}{27}$ de los azares. La suma de los azares que hacen ganar al primer jugador es entonces $\frac{1}{3} + \frac{2}{9} + \frac{2}{27}$, lo que hace $\frac{17}{27}$. Así, hacía ver Fermat que la extensión ficticia a un cierto número de partidas no es otra cosa que la reducción de las diversas fracciones a un mismo denominador.

El razonamiento de Fermat puede escribirse de la siguiente manera:

Consideremos los eventos siguientes:

A: el primer jugador gana el juego.

A_1 : el primer jugador gana la siguiente partida.

²Implícitamente, esto es lo que hace ver Galileo en el problema particular que se planteaba con relación al lanzamiento de 3 dados.

A_2 : Alguno de los otros dos jugadores gana la siguiente partida y la que sigue a ésta la gana el primer jugador.

A_3 : en las siguientes dos partidas, cada uno de los otros dos jugadores gana una y la siguiente la gana el primer jugador.

Entonces $A = A_1 \cup A_2 \cup A_3$ y, como los eventos A_1 , A_2 y A_3 son mutuamente excluyentes, se tiene $P(A) = P(A_1) + P(A_2) + P(A_3)$.

De esta manera, se ve que lo que hizo Fermat fue introducir implícitamente la regla de la suma para eventos mutuamente excluyentes.

Cuando Pascal conoció el método usado por Fermat en el problema de las partidas comentó que este método es el primero que le vino a la mente para solucionar este problema pero que, como el trabajo con las combinaciones es excesivo, buscó una abreviación, encontrando propiamente otro método. Este comentario parece importante pues deja ver que era conocido y aceptado que para resolver problemas de azar podía recurrirse a contar las distintas posibilidades y de ahí las que son favorables al evento deseado.

Pasemos entonces al método utilizado por Pascal en el problema de las partidas. Tomemos el mismo caso de dos jugadores que ya consideramos, en el que al primero, P, le faltan dos partidas para ganar y al segundo, Q, tres. Para encontrar la forma en que deben repartirse las apuestas si el juego se detiene en estas condiciones, Pascal siguió un método recursivo. En el caso que estamos considerado comenzaba por establecer la forma de la repartición cuando al primer jugador le falta una partida para ganar y al segundo dos. Supongamos que el total de las apuestas es 64, entonces, en la situación dada, al jugar la siguiente partida hay dos posibilidades, la primera es que P la gane, en cuyo caso gana el juego y por lo tanto toda la apuesta, la segunda es que Q la gane en cuyo caso P y Q quedan en igualdad de condiciones y debe entonces tocar a cada uno la mitad de las apuestas, es decir 32. Entonces en un caso a P le tocan 64 y en otro 32, así que, cualquiera que sea el caso, P tiene asegurado 32 y los otros 32 de las apuestas pueden corresponder a P o a Q con un azar igual; por lo tanto, de esos 32 la mitad debe ser para P y la otra para Q. Es decir, cuando a P le falta un punto y a Q dos, a P le corresponde $32 + 26 = 48$ y a Q 16.

A partir de este caso, Pascal pasaba a otro en el que a P le falta un punto y a Q tres. En esta situación, si se juega la siguiente partida, P puede ganar toda apuesta o bien 48 por el primer caso. Por lo tanto a P le corresponde $48 + \frac{1}{2}(16) = 56$ y a Q 8.

En seguida podía pasar al caso en que a P le faltan dos puntos y a Q tres. En esta situación, si se juega la siguiente partida, P puede quedar faltándole un punto y tres a Q, en cuyo caso le corresponde 56 por el segundo caso; o bien, si Q gana

esta partida, quedan en igualdad de circunstancias y toca a cada uno 32. Entonces P tiene asegurados 32 y puede ganar $56 - 32 = 24$ con un azar igual que Q; así que entonces a P le corresponde $32 + \frac{1}{2}(24) = 44$ y a Q $8 + \frac{1}{2}(24) = 20$.

La proporción $44 : 20$ encontrada por Pascal es, efectivamente, la misma que la proporción $11 : 5$ encontrada con el método de Fermat.

Pascal no se contentaba con encontrar la solución a un problema particular, su búsqueda fue siempre en el sentido de encontrar reglas generales que le permitieran encontrar soluciones en una diversidad de situaciones. Para ejemplificar esto observemos primero que, siguiendo el método de Pascal, se obtienen los siguientes resultados: Cuando a P le falta 1 punto y a Q 4 puntos, a P le corresponde 60 de la apuesta; cuando a P le faltan 2 puntos y a Q 4, a P le corresponde 52 de la apuesta y, finalmente, cuando a P le faltan 3 puntos y a Q 4, a P le corresponde 42 de la apuesta. Entonces, decía Pascal, cuando P ha ganado 1 punto y Q ninguno, le tocan a P 42, es decir P tiene derecho a 10 de la apuesta de Q; en otras palabras, el valor, sobre la apuesta del contrario, de la primera partida es 10; cuando P ha ganado 2 puntos y Q ninguno, le tocan a P 52, es decir P tiene derecho a 20 de la apuesta de Q; restándole a esta cantidad el valor de la primera partida resulta que el valor de la segunda partida es 10. De la misma manera se determina que el valor de la tercera partida es 8 y el de la cuarta 4. Pascal entonces observa que dado cualquier número N , cuando se juega a N partidas, el valor de la primera partida es igual al de la segunda y el valor de la última partida es igual a la mitad del valor de la última partida cuando se juega a $N - 1$ puntos; así, si la apuesta de cada uno es 32, cuando se juega a un punto el valor de la última partida es 32; cuando es a 2, 16; cuando es a 3, 8; etc.

Pascal fundamentó y amplió las bases de su método en su tratado sobre el triángulo aritmético. Ahí, dice Pascal que su método está basado en los siguientes dos principios:

1o. Si uno de los jugadores se encuentre en una situación tal que, independientemente de que gane o pierda, una suma le debe pertenecer sin que al azar pueda quitársela, entonces no debe repartirla, sino tomarla entera, ya que la repartición debe ser proporcional al azar y como no hay azar de perderla, debe retirar todo sin repartirlo.

2o. Si dos jugadores se encuentran en una situación tal que, si uno gana le pertenecerá una cierta suma y si pierde, ésta pertenecerá al otro; si el juego es de puro azar y hay tantos azares para uno como para el otro y quieren retirarse sin jugar y tomar lo que legítimamente les pertenece, entonces deben partir la suma que está al azar por la mitad y tomar cada uno una parte.

De estos dos principios, Pascal obtiene dos corolarios, que son esencialmente el mismo y puede darse de la siguiente forma:

COROLARIO 6.1. *Si dos jugadores juegan un juego de puro azar con la condición de que si el primero gana se le dará una cierta suma A y si pierde se le dará una suma B . Entonces, si quieren retirarse sin jugar y tomar cada uno lo que les pertenece, el primero debe tomar $\frac{1}{2}(A + B)$.*

Los dos principios son intuitivamente evidentes, no así el Corolario, aunque éste se obtenga en forma inmediata de los principios. Este corolario encierra un concepto de gran importancia en la teoría de la Probabilidad, el de Esperanza. En un juego de azar, este concepto corresponde precisamente a lo que debe recibir un jugador en caso de que decida no jugar el juego, lo cual puede entonces interpretarse como una estimación, hecha antes de realizar el juego, de lo que recibirá el jugador.

Aplicando repetidamente el Corolario, Pascal podía resolver el problema de las partidas en cualquier circunstancia. Pero Pascal hizo aún más, pues encontró que hay una relación entre las soluciones al problema de las partidas y el triángulo aritmético. Esta relación es la que estudiamos en el capítulo 3 del primer volumen de este libro y que demostramos utilizando básicamente el método de Fermat. Es interesante la demostración que hizo Pascal de esta relación pues no se basa en el método de Fermat sino en una aplicación directa de su corolario.

El método de Pascal es esencialmente el mismo que utiliza más tarde Huygens para resolver tanto el problema de las partidas como el problema de los dados, solo que, como veremos, el método de Huygens, a pesar de sus limitaciones, es más general que el de Pascal. Pasemos a analizar este método.

Para Huygens, el tipo de problemas planteados eran de gran importancia, pues consideraba que su estudio era la base de una nueva materia.

Huygens comienza su trabajo diciendo que, si bien en los juegos de azar los resultados son inciertos, la posibilidad que un jugador tiene de ganar o de perder tiene sin embargo un valor determinado. Para encontrar este valor, Huygens partía de una hipótesis en la cual introduce el concepto de juego justo. Esta hipótesis se enuncia como sigue:

HIPÓTESIS 6.2. *En un juego, la posibilidad que se tiene de ganar alguna cosa tiene un valor tal que si se posee este valor se puede uno procurar la misma posibilidad en un juego justo.*

Por un juego justo, Huygens entendía un juego entre un número cualquiera de jugadores en el cual, o bien todos los jugadores tienen la misma posibilidad de ganar cierta cantidad, o bien cada uno de los jugadores tiene igual número de posibilidades de ganar cierta cantidad que de perderla.

La hipótesis entonces significa lo siguiente: supongamos que un jugador P participa en un juego en el que puede recibir cantidades A_1, \dots, A_n , dependiendo del resultado del juego; supongamos, para facilitar el razonamiento, que todas esas cantidades son positivas, es decir, el jugador P gana siempre alguna cantidad cualquiera que sea el resultado, lo que varía es el monto de la cantidad. Este juego es, desde luego, favorable al jugador P pues no arriesga nada. Se quiere estimar, antes de realizar el juego, el valor x que recibe P. Este valor x puede interpretarse, como ya vimos en el método de Pascal, como lo que pertenece a P en caso de que decida no jugar el juego; entonces podemos decir también que para P el juego tiene un valor x . Esto significa que si, para jugar el juego, P paga una cantidad x , entonces el juego no será ni favorable ni desfavorable para P, ni tampoco para el que le paga a P, es decir, el juego será un juego justo. La hipótesis de Huygens es entonces intuitivamente evidente, dice simplemente que el valor de un juego para una persona que participa en él es tal que si esa persona paga ese valor por jugar el juego, entonces el juego no es favorable a ninguno de los jugadores, es decir, es un juego justo. Incluso la hipótesis puede tomarse como una definición del valor de un juego en términos de un juego justo; la definición se interpretaría, intuitivamente, diciendo que el valor de un juego así definido es una estimación, realizada antes de jugar el juego, de lo que la persona recibirá.

De su hipótesis, Huygens dedujo tres proposiciones, las cuales serían las únicas que usaría para resolver todos los problemas que se planteó. Estas son las siguientes:

PROPOSICIÓN 6.3. *Tener iguales posibilidades de obtener a o b me vale $\frac{a+b}{2}$*

PROPOSICIÓN 6.4. *Tener iguales posibilidades de obtener a , b o c me vale $\frac{a+b+c}{3}$*

PROPOSICIÓN 6.5. *Tener r posibilidades de obtener a y s posibilidades de obtener b , las posibilidades siendo equivalentes, me vale $\frac{ra+sb}{r+s}$*

Como puede verse, la proposición 6.3 es exactamente el Corolario de Pascal; mientras que las proposiciones 6.4 y 6.5 generalizan la proposición 6.3, con la número 2 Huygens resuelve el problema de las partidas para el caso de 3 jugadores. Nótese que Huygens especifica en la proposición 6.5 que las posibilidades deben ser equivalentes (i. e. equiprobables).

Para ilustrar el método de demostración de estas proposiciones se expone a continuación la de la número 3:

Sea x el valor del juego y consideremos un nuevo juego definido de la siguiente manera: juego contra otras dos personas teniendo cada una las mismas posibilidades de ganar y apostando x cada una, conviniendo además con la primera que si gana ella me da una cantidad b y si yo gano le doy una cantidad b y con la segunda que si gana ella me da una cantidad c y si yo gano le doy una cantidad c . Evidentemente, este nuevo

juego así definido es un juego justo. En este nuevo juego, si gano recibo $3x - b - c$, si gana la primera persona recibo b y si gana la segunda recibo c ; es decir, tengo iguales posibilidades de obtener $3x - b - c$, b o c . Haciendo entonces $3x - b - c = a$ el nuevo juego resulta equivalente al que teníamos originalmente y obtenemos $x = \frac{a+b+c}{3}$.

Haciendo uso de sus proposiciones, Huygens resolvió el problema de las partidas para algunos casos particulares. Cuando se trata de dos jugadores utilizó la proposición 6.3 siguiendo esencialmente el mismo método que Pascal, es decir, para llegar a una cierta situación comenzaba por considerar otras situaciones más simples. Cuando se trata de 3 jugadores Huygens usó la proposición 6.4, ésta no aparece en el trabajo de Pascal, sin embargo él afirma que su método sirve también para el caso de 3 jugadores, lo cual no es evidente si no se tiene una proposición del tipo de la 6.4. Inténtese por ejemplo resolver el caso más simple de 3 jugadores usando el método de Pascal, es decir, supongamos que 3 personas están jugando a cierto número de partidas y que a la primera le falta una partida para ganar, una también a la segunda y dos a la tercera. Utilizando la proposición 6.4 diríamos:

Si el juego continúa y la siguiente partida es ganada por la primera persona, entonces a esta le toca toda la apuesta, llamémosla A y a las otras no les toca nada; si es la segunda persona la que la gana, entonces a ésta le toca A y a las otras nada; finalmente, si es la tercera persona la que la gana entonces quedan en igualdad de circunstancias, tocándole a cada una $\frac{1}{3}A$. Por lo tanto, si el juego se detiene en las condiciones dadas inicialmente, a la primera persona le corresponde, por la proposición 6.4, $\frac{A+0+\frac{1}{3}A}{3} = \frac{4}{9}A$; a la segunda lo mismo y a la tercera $\frac{0+0+\frac{1}{3}A}{3} = \frac{1}{9}A$.

La proposición 6.5 se puede utilizar para resolver el problema de las partidas en el caso en que cada jugador tenga distintas posibilidades de ganar cada partida, sin embargo Huygens no hizo esta aplicación. También la proposición 6.5 es inmediatamente generalizable al caso de cualquier número n de cantidades a_1, \dots, a_n con posibilidades iguales a r_1, \dots, r_n , respectivamente, en cuyo caso el valor del juego resultaría $\frac{r_1 a_1 + \dots + r_n a_n}{r_1 + \dots + r_n}$; esta generalización tampoco la hizo Huygens.

Observemos por último que el método de Huygens y, por lo tanto, el de Pascal involucran implícitamente el concepto de Esperanza Condicional y, con éste, el de Probabilidad Condicional. Consideremos por ejemplo un juego a cuatro puntos que se detiene cuando P ha ganado dos puntos y Q uno. Nos preguntamos entonces por la proporción de las apuestas que corresponde a P, o bien, partiendo de la situación dada, por el valor del juego para P.

Sea a el total de las apuestas y x el valor del juego para P partiendo de la situación dada. Si supiéramos cuantas posibilidades tienen P y Q de ganar, partiendo de la situación dada, obtendríamos fácilmente lo que le corresponde a cada uno, pues si r

es el primer número y s el segundo, entonces, por la proposición 6.5, al jugador P le corresponde $\frac{ra}{r+s}$ y al jugador Q $\frac{sa}{r+s}$. Es decir, el valor del juego para P es $x = \frac{ra}{r+s}$.

Como r y s no se conocen, lo que hacemos es, partiendo de la situación dada, suponer que se juega la siguiente partida, entonces llamando x_1 al valor del juego para P cuando P gana la siguiente partida y x_2 su valor cuando Q gana la siguiente partida, usando la proposición 1, obtenemos $x = \frac{x_1+x_2}{2}$.

Utilizando la terminología moderna, estamos considerando los eventos:

A_1 : El jugador P gana la siguiente partida.

B_1 : El jugador Q gana la siguiente partida.

Entonces, x es el valor del juego para P (i.e. el valor que espera recibir P), mientras que x_1 y x_2 son los valores del juego para P condicionados a la ocurrencia de A_1 y B_1 respectivamente (i.e. esperanzas condicionales). De manera que si llamamos X a lo que recibe P en este juego, la fórmula $x = \frac{x_1+x_2}{2}$ puede escribirse en la forma $E[X] = E[X | A_1]P(A_1) + E[X | B_1]P(B_1)$, la cual podríamos llamar la regla de la probabilidad total para esperanzas.

La fórmula en términos de esperanzas puede escribirse en términos de probabilidades. En efecto, si llamamos r_1 a las posibilidades que P tiene de ganar el juego cuando A_1 ocurre y s_1 a las posibilidades de Q, entonces $x_1 = \frac{r_1a}{r_1+s_1}$. Pero si llamamos A al evento consistente en que P gane el juego, el cociente $\frac{r_1}{r_1+s_1}$ es precisamente la probabilidad condicional de A dada la ocurrencia de A_1 , es decir $x_1 = aP(A | A_1)$. De la misma manera, se tiene $x_2 = aP(A | B_1)$ y $x = aP(A)$. Por lo tanto la fórmula $E[X] = E[X | A_1]P(A_1) + E[X | B_1]P(B_1)$ puede escribirse en la forma $aP(A) = aP(A | A_1)P(A_1) + aP(A | B_1)P(B_1)$ o, lo que es lo mismo, $P(A) = P(A | A_1)P(A_1) + P(A | B_1)P(B_1)$, que no es otra cosa que la regla de la probabilidad total. Nótese que esta regla se obtiene aquí sin necesidad de usar la regla del producto, la cual, por otro lado, también puede obtenerse siguiendo el método de Huygens.

Los diferentes métodos utilizados por Fermat-Pascal-Huygens para resolver el problema de las partidas encierran, como hemos visto, varios conceptos y resultados importantes en la teoría de la Probabilidad. Está ahí el concepto de probabilidad a la manera clásica como un cociente de casos favorables y total de casos, haciéndose ver que el total de casos deben ser equiprobables; está también el concepto de esperanza y junto con éste el de juego justo; están asimismo los conceptos de esperanza y de probabilidad condicional, así como la regla de la suma para el caso de eventos mutuamente excluyentes y la regla de la probabilidad total.

Estas consideraciones pueden servir para ubicar el trabajo de Pascal-Fermat-Huygens, pues en conjunto encierran una gran riqueza, de hecho todas las bases para el desarrollo del Cálculo de Probabilidades clásico, incluyendo su parte teórica. No son ellos, claro está, los que abstraen los conceptos y resultados anteriores de los problemas particulares que tratan, pero sí son estos problemas los que servirán de base para la sistematización que más tarde realizará Jacques Bernoulli por un lado y Abraham de Moivre por el otro.

Con relación al problema de las partidas, tanto Pascal como Fermat y Huygens hicieron aportaciones. Parece ser que las mayores pertenecen a Fermat y a Huygens, pues es de las consideraciones que ellos hacen que se pueden abstraer los conceptos y resultados mencionados más arriba.

6.3.2. Problemas con dados. En la correspondencia entre Pascal y Fermat se hace referencia a dos problemas de dados propuestos por de Meré, éstos son los problemas 2 y 3.

Pascal y Fermat dieron mayor atención al problema de las partidas que a estos dos problemas; sin embargo, veremos que las consideraciones que de ellos se derivan encierran tanta o más riqueza que el problema de las partidas.

Los métodos que usaron Pascal y Fermat para resolver estos dos problemas son, como ya dijimos, desconocidos; lo que es cierto es que no están basados en una aplicación directa de la definición clásica de probabilidad, excepto quizá el usado para el problema 1. Este fue resuelto no sólo por Pascal y Fermat sino por Roberval y el mismo de Méré, lo cual podría indicarnos que para su solución fue usado el método, aparentemente aceptado y usado en esa época, consistente en determinar todas las posibilidades en los lanzamientos y de ahí aquellas que son favorables al evento deseado. El conocimiento de todas las posibles formas en que pueden caer cualquier número de dados, problema que estaba ya resuelto, facilitaba esta tarea pues ya solo era necesario encontrar de ahí los casos favorables al evento en consideración. En el caso del problema 2 la solución es $n = 4$, así que comenzando por el caso $n = 1$ el proceso para llegar a la solución no resulta ni muy largo ni muy complejo, en cambio, la respuesta $n = 25$ para el problema 3 hace que ese mismo proceso se alargue y se complique.

Pascal y Fermat encontraron las soluciones correctas a los dos problemas anteriores. Por cierto que el resultado sorprendía a de Méré pues decía que si con un dado, en cual tiene 6 caras, se requieren 4 lanzamientos para obtener un determinado valor, por qué con dos dados, los cuales tienen 36 caras, y dado que 4 es a 6 como 24 es a 36, no se requieren 24 lanzamientos para obtener un determinado valor.

El único problema con dados que fue resuelto en la correspondencia que se conoce entre Pascal y Fermat es uno que resuelve Fermat en una carta a Pascal a propósito de un problema que este último le había planteado.

PROBLEMA 7. *Si se trata de obtener por lo menos un 6 en ocho lanzamientos de un dado y se ha jugado ya 3 veces sin obtenerlo y el contrario propone que no se haga el 4o. lanzamiento (pero si los siguientes), ¿Cuánto le pertenece al que está tirando por dejar de hacerlo?*

Antes de resolver este problema, Fermat dio el siguiente argumento: Si trato de hacer un 6 en ocho lanzamientos de un dado y, después que el dinero está en juego, convenimos en que no haré el primer lanzamiento, entonces es necesario, *por mi principio*, que saque del juego un sexto del total. Si después de eso, convenimos en que no haré el segundo lanzamiento, entonces debo sacar un sexto de lo restante que es $\frac{5}{36}$ del total; y si después de eso convenimos en que no haré el tercer lanzamiento, entonces debo sacar un sexto de lo restante, que es $\frac{25}{216}$ del total; si todavía se conviene en que no haga el cuarto lanzamiento debo sacar un sexto de lo restante, que es $\frac{125}{1296}$ del total.

Pero, agregaba Fermat, en el problema planteado, como no se obtuvo nada en las 3 primeras tiradas, la suma total todavía está en juego, de manera que si se conviene que no se haga el 4o. lanzamiento, el que tira debe tomar como indemnización un sexto del total.

Se desconoce a que principio se refiere Fermat en su argumentación; es posible que se refiera a uno según el cual, si todos los azares son iguales (i.e. equiprobables) y hay r casos favorables para obtener cierta cantidad x de un total de $r + s$ casos, entonces lo que corresponde al jugador antes de realizarse el juego es $\frac{r}{r+s}x$. Como puede verse, este principio es un caso particular de la proposición 3 de Huygens, aunque su demostración puede darse siguiendo las ideas de Fermat dadas en su solución al problema de las partidas, ahí él divide las apuestas en la misma proporción que los casos favorables a cada jugador, de ahí se sigue que si r son los favorables al primero y s al segundo, entonces al primero corresponde $\frac{r}{r+s}$ de las apuestas y al segundo $\frac{s}{r+s}$ de las apuestas.

Aplicando este principio en cada paso de su argumento se obtiene la solución que él da. Obsérvese que lo que queda después de sacar la primera cantidad es $\frac{5}{6}$ del total, lo que queda después de sacar la segunda es $\frac{25}{36}$ del total, después de sacar la tercera $\frac{125}{216}$ del total y ésta es la forma en que Fermat escribe estas cantidades. La regla de aquí es evidente: lo que queda después de sacar la n ésima cantidad es $\left(\frac{5}{6}\right)^n$.

Del argumento de Fermat puede inferirse que su método de solución al problema 2 de dados es el mismo que usó ahí; es decir, si en n lanzamientos, el primero vale $\frac{1}{6}$ del total, el segundo $\frac{1}{6}$ de lo que resta y así sucesivamente, entonces el valor de todos los lanzamientos juntos es $\frac{1}{6} + \frac{1}{6}\frac{5}{6} + \frac{1}{6}\frac{25}{36} + \frac{1}{6}\frac{125}{216} + \dots$ del total.

El problema se reduce a encontrar entonces un número n para el cual la suma anterior sea ya mayor que $\frac{1}{2}$. Haciendo paso a paso esta suma se obtiene que ya la suma de los 4 primeros términos es mayor que $\frac{1}{2}$. No se sabe si Fermat siguió este procedimiento para encontrar $n = 4$ o si encontró la expresión general de la suma para cualquier n , ésta es fácil de obtener pues la suma es una progresión:

$$\frac{1}{6} + \frac{1}{6} \left(\frac{5}{6}\right) + \dots + \frac{1}{6} \left(\frac{5}{6}\right)^{n-1} = 1 - \left(\frac{5}{6}\right)^n$$

El problema se reduce entonces a encontrar el más pequeño n para el cual $\left(\frac{5}{6}\right)^n < \frac{1}{2}$.

Obsérvese que en el método de Fermat está implícita la regla de la suma para eventos mutuamente excluyentes pues el j -ésimo término de la suma $\frac{1}{6} + \frac{1}{6} \left(\frac{5}{6}\right) + \dots + \frac{1}{6} \left(\frac{5}{6}\right)^{n-1}$ corresponde a la probabilidad de obtener el primer seis exactamente en el j -ésimo lanzamiento.

Huygens prestó mayor atención a los problemas con dados. En su libro se encuentran resueltos los problemas 2 y 3 y algunos otros que planteamos a continuación:

El estudio de problemas con dados lo comenzó Huygens haciendo ver que con un dado se pueden obtener 6 resultados diferentes, todos igualmente posibles porque se supone que el dado tiene la forma de un cubo perfecto. Como puede verse, aquí está la idea de equiprobabilidad. En seguida hacía ver que con dos dados se pueden obtener $6 \times 6 = 36$ resultados diferentes, también todos igualmente posibles; con tres dados se tienen $36 \times 6 = 216$ y así sucesivamente. Por último, encontró el número de resultados, de entre todos los posibles, con los cuales se obtiene cada una de las diferentes sumas que pueden obtenerse con dos y tres dados.

Después de estas consideraciones pasaba ya a la solución de los diferentes problemas. Su método de solución es esencialmente el mismo que usó para el problema de las partidas, basado en sus 3 proposiciones.

Para resolver el problema 2, decía Huygens: El jugador que acepta obtener un 6 con un solo lanzamiento tiene 1 posibilidad de ganar y 5 de perder, así que, llamando A a la apuesta, tiene 1 posibilidad de obtener A y 5 de no obtener nada, lo que le vale $\frac{1}{6}A$ por la proposición 6.5. El jugador que acepta obtener un 6 en dos lanzamientos tiene 1 posibilidad de obtenerlo en el primer lanzamiento y 5 de no obtenerlo; si lo obtiene en el primer lanzamiento, gana A , si no lo obtiene, todavía puede hacerlo en el segundo lanzamiento, lo cual le vale $\frac{1}{6}A$ por la primera parte de la demostración;

por lo tanto, al comienzo tiene 1 posibilidad de obtener A y 5 de obtener $\frac{1}{6}A$, lo cual le vale $\frac{A+5(\frac{1}{6}A)}{6} = \frac{11}{36}A$ por la proposición 6.5. En los pasos siguientes el razonamiento es similar: con tres lanzamientos, si obtiene 6 en el primero gana A , si no lo obtiene tiene derecho a $\frac{11}{36}A$ por el segundo paso; por lo tanto, al comienzo le corresponde $\frac{A+5(\frac{11}{36}A)}{6} = \frac{91}{216}$ por la proposición 6.5. Con 4 lanzamientos se obtiene $\frac{671}{1296}$, es decir, más de la mitad de A .

El método de solución consiste entonces, como en el problema de la división de apuestas, en ir del caso más simple al más complejo con un procedimiento recursivo. En este problema el método de Huygens es también esencialmente un método basado en probabilidades condicionales. Utilizando la terminología moderna y planteado en términos de probabilidades, el método es esencialmente el siguiente:

Consideremos los siguientes eventos:

A_n : Se obtiene por lo menos un 6 en n lanzamientos.

B_1 : Se obtiene 6 en el primer lanzamiento.

Entonces:

$$\begin{aligned} P(A_n) &= P(A_n | B_1)P(B_1) + P(A_n | \bar{B}_1)P(\bar{B}_1) = \frac{1}{6} + \frac{5}{6}P(A_n | \bar{B}_1) \\ &= \frac{1}{6} + \frac{5}{6}P(A_{n-1}) = \frac{1}{6} + \frac{5}{6} \left[\frac{1}{6} + \frac{5}{6}P(A_{n-2}) \right] = \frac{1}{6} + \frac{1}{6} \left(\frac{5}{6} \right) + \left(\frac{5}{6} \right)^2 P(A_{n-2}) \\ &= \frac{1}{6} + \frac{1}{6} \left(\frac{5}{6} \right) + \left(\frac{5}{6} \right)^2 \left[\frac{1}{6} + \frac{5}{6}P(A_{n-3}) \right] \\ &= \frac{1}{6} + \frac{1}{6} \left(\frac{5}{6} \right) + \frac{1}{6} \left(\frac{5}{6} \right)^2 + \left(\frac{5}{6} \right)^3 P(A_{n-3}) \\ &= \dots = \frac{1}{6} + \frac{1}{6} \left(\frac{5}{6} \right) + \frac{1}{6} \left(\frac{5}{6} \right)^2 + \dots + \frac{1}{6} \left(\frac{5}{6} \right)^{n-1} \end{aligned}$$

Expresión que coincide con la que se puede encontrar con el método de Fermat.

Obsérvese, en particular, que en el método de solución está implícita la regla de la probabilidad total. El método está basado, como en el caso del problema de las partidas, directamente en el concepto de esperanza y de juego justo. Es decir, Huygens no calculaba directamente probabilidades sino lo que el jugador espera recibir en cada lanzamiento, en donde lo que espera recibir está definido en función de un juego justo.

Para resolver el problema 3 Huygens siguió el mismo método que utilizó para el problema 2, sin embargo la solución $n = 25$ lo hace relativamente largo, Huygens encontró sin embargo una manera de simplificarlo. Decía Huygens:

El que juega a un solo lanzamiento (de dos dados) tiene 1 posibilidad de ganar A y 35 posibilidades de no ganar nada, así que le corresponde $\frac{1}{36}A$. El que juega a dos lanzamientos tiene 1 posibilidad de ganar A y 35 posibilidades de obtener $\frac{1}{36}A$ (por el primer paso), lo que le vale $\frac{A+35(\frac{1}{36}A)}{36} = \frac{71}{1296}A$. El que juega a cuatro lanzamientos, gana A si obtiene par de seises en los primeros dos lanzamientos, si no lo obtiene gana, por el segundo paso, $\frac{71}{1296}A$; pero, también por el segundo paso, hay 71 posibilidades de obtener par de seises en los dos primeros lanzamientos y $1296-71 = 1225$ posibilidades de no obtenerlo; por lo tanto, en 4 lanzamientos le corresponde $\frac{71A+1225(\frac{71}{1296}A)}{1296}$, es decir $\frac{178991}{1679616}A$.

De aquí calculaba lo que le corresponde en 8 lanzamientos, luego en 16 lanzamientos y, usando estos dos casos, encuentre lo que le corresponde en 24 lanzamientos. Finalmente encontró la solución correcta, es decir que el que juega a 24 lanzamientos tiene todavía una ligera desventaja y que se puede aceptar la partida con ventaja jugando a 25 lanzamientos por lo menos.

La simplificación que hacía Huygens está basada en la siguiente consideración:

Tomemos el caso del que juega a obtener par de seises en 2 lanzamientos, entonces llamando x a lo que corresponde el jugador en este caso y A_1 al evento consistente en obtener par de seises en el primer lanzamiento, ya hemos visto que Huygens calculaba x de la siguiente manera:

$$x = x_1P(A_1) + x_2P(\bar{A}_1)$$

en donde x_1 es lo que gana cuando ocurre A_1 y x_2 lo que gana cuando no ocurre A_1 .

Lo que se obtiene como valor de x es una fracción $\frac{x}{A}$ multiplicada por la apuesta A . Lo que decía Huygens es que la fracción es un cociente de resultados favorables entre total de resultados, es decir, es una probabilidad. Efectivamente, hemos visto que llamando B al evento consistente en obtener par de seises en dos lanzamientos, podemos escribir $x = AP(B)$. Es decir, en el razonamiento de Huygens está implícita la idea de que sus proposiciones 6.3, 6.4 y 6.5 pueden expresarse no solo en términos de esperanzas sino también en términos de probabilidades. En otras palabras, dentro del concepto de Esperanza que definió Huygens está contenido implícitamente el concepto clásico de probabilidad. En otras palabras, de aquí se refuerza la idea que hemos dado ya en el sentido de que dentro del concepto de esperanza que definió Huygens está contenido implícitamente el concepto clásico de probabilidad. La última idea de Huygens expresa lo siguiente: si B es un evento y con la ocurrencia de B un jugador recibe una cantidad A y con la no ocurrencia de B no recibe nada, entonces, llamando x a lo que corresponde al jugador al inicio del juego, se tiene $x = AP(B)$, e inversamente, si lo que se conoce es x entonces $\frac{x}{A} = P(B)$.

En este problema, al igual que en el anterior, se puede encontrar una expresión simple para lo que corresponde a un jugador que acepta la obtención de par de seises en n lanzamientos de un par de dados, llamando x_n a esa cantidad, se obtendría $x_n = [1 - (\frac{35}{36})^n] A$, en donde A es el total de las apuestas.

Este mismo resultado puede obtenerse de una manera más simple utilizando el mismo método de Huygens considerando lo que lo que corresponde al contrario en lugar de lo que corresponde al jugador. Este cálculo es como sigue:

En un lanzamiento, el contrario tiene 35 posibilidades de obtener A y 1 posibilidad de no obtener nada; le corresponde entonces $\frac{35}{36}A$. En dos lanzamientos tiene 1 posibilidad de no obtener nada (si el jugador obtiene par de seises en el primer lanzamiento) y 35 posibilidades de obtener $\frac{35}{36}A$ por el primer paso; le corresponde entonces $\frac{35(\frac{35}{36}A)}{36} = (\frac{35}{36})^2 A$. Con el mismo razonamiento se encuentre entonces que en 3 lanzamientos le corresponde $(\frac{35}{36})^3 A$; en 4 lanzamientos $(\frac{35}{36})^4 A$ y así sucesivamente; es decir, en n lanzamientos le corresponde $(\frac{35}{36})^n A$. Por lo tanto, al jugador que trata de obtener par de seises en n lanzamientos le corresponde $A - (\frac{35}{36})^n A$.

Este hecho muestra que no siempre la solución más simple es la primera que se ocurre e incluso puede no ser evidente; Fermat por ejemplo tampoco encontró esta forma simple de la solución al problema planteado, aun cuando él calculaba probabilidades en casos análogos usando la definición clásica de probabilidad; en este caso, de $(36)^n$ posibles resultados equiprobables hay $(35)^n$ desfavorables y resulta entonces inmediato que un juego a n lanzamientos, en el que se trate de obtener par de seises, vale $[1 - (\frac{35}{36})^n] A$. Lo inmediato o simple de una solución a un problema requiere pues, a veces, de ensayos de solución y de maduración de determinados conceptos.

Si se analizan las soluciones de Huygens y de Fermat a los problemas con dados, se verá que éstas tienen implícito el uso de la independencia de los lanzamientos. Sin embargo, la independencia de experimentos o, más generalmente, de eventos, es un concepto que se clarificó hasta más tarde. Los problemas con dados jugaron un papel importante en este proceso pues ilustraban el concepto ampliamente. Huygens mismo considera otro problema con dados, el cual, al ser generalizado por Jacques Bernoulli, adquiriría una importancia singular en la Teoría de la Probabilidad, nos referimos al problema 4, el cual es equivalente a encontrar en cuántos lanzamientos de un dado se puede contar con obtener dos veces un 6 (por lo menos). La solución a este problema usando el método de Huygens es simple, aunque algo laboriosa pues se requiere encontrar primero lo que vale el juego para el caso de 2 lanzamientos, luego para el caso de 3, 4, 5, ..., etc., hasta el momento en que el juego ya resulte favorable, se encuentre que se puede jugar con ventaja a obtener 2 series (por lo menos) con un dado en 10 lanzamientos.

Con relación a este problema, Jacques Bernoulli más tarde encontraría las probabilidades tipo binomial ([1]); de manera específica, encontraría que la probabilidad de obtener exactamente k seises en n lanzamientos de un dado es $\binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$. Este resultado es de fundamental importancia en su trabajo pues con él se puede calcular la probabilidad de obtener una frecuencia de seises igual a $\frac{k}{n}$ en n lanzamientos de un dado y de aquí encontrar una relación entre la frecuencia de ocurrencia de un evento y su probabilidad, para obtener lo que se llama el Teorema de Bernoulli.

Así, las soluciones de Fermat y Huygens a los problemas con dados que se plantearon contienen implícitamente los mismos conceptos y resultados que los contenidos en las soluciones al problema de las partidas; pero, además, la posibilidad de repetir los lanzamientos ilustra un concepto también de importancia, el de independencia y, por otra parte, el problema 4 que plantea Huygens daría origen a un tipo de probabilidades de gran importancia en la Teoría de la Probabilidad, las de tipo binomial.

Huygens resolvió en su libro algunos otros problemas, entre los que destaca el problema 5. La importancia de ese problema radica en que se refiere a un experimento el cual admite una infinidad de posibles resultados, rebasando el marco de la misma definición clásica de probabilidad. La solución de Huygens es como sigue:

Sea x el valor del juego para Q y a el total de las apuestas. El valor del juego para P es entonces $a - x$. Sea además y el valor del juego para Q cuando sea su turno de lanzar los dados. Al iniciarse el juego, Q tiene 5 posibilidades de obtener 0 y 31 posibilidades de obtener y , por lo tanto, $x = \frac{31}{36}y$. Por otra parte, cada vez que Q tenga el turno para lanzar los dados, tiene 6 posibilidades de obtener a y 30 de obtener x , por lo tanto, $y = \frac{6a+30x}{36}$. Resolviendo el sistema de ecuaciones se obtiene $x = \frac{31}{61}a$, de manera que los valores del juego para P y Q, respectivamente, están en la proporción 30 : 31.

Nuevamente, la solución de Huygens contiene implícitamente el uso de probabilidades condicionales. La solución moderna podría plantearse definiendo los siguientes eventos:

C : El juego es ganado por el jugador Q.

A_1 : Se obtiene éxito en el primer lanzamiento.

B_1 : Se obtiene fracaso en el primer lanzamiento.

A_2 : Se obtiene éxito en el segundo lanzamiento.

B_2 : Se obtiene fracaso en el segundo lanzamiento.

para los cuales se tiene:

$$\begin{aligned}
P(C) &= P(C | A_1)P(A_1) + P(C | B_1)P(B_1) = P(C | B_1)P(B_1) = \frac{31}{36}P(C | B_1) \\
&= \frac{31}{36} [P(C | B_1 \cap A_2)P(A_2) + P(C | B_1 \cap B_2)P(B_2)] \\
&= \frac{31}{36} [P(A_2) + P(C)P(B_2)] = \frac{31}{36} \left[\frac{6}{36} + \frac{30}{36}P(C) \right]
\end{aligned}$$

Así que $P(C) = \frac{31}{61}$.

O, de manera equivalente, definiendo los eventos:

C : El juego es ganado por el jugador Q.

A_1 : Se obtiene éxito en el primer lanzamiento.

E : Se obtiene fracaso en el primer lanzamiento y éxito en el segundo.

F : Se obtiene fracaso en los primeros dos lanzamientos.

para los cuales se tiene:

$$\begin{aligned}
P(C) &= P(C | A_1)P(A_1) + P(C | E)P(E) + P(C | F)P(F) \\
&= P(E) + P(C)P(F) = \frac{31}{36} \frac{6}{36} + \frac{31}{36} \frac{30}{36} P(C)
\end{aligned}$$

El libro de Huygens termina con el planteamiento de algunos problemas, entre los cuales destaca el problema 6³.

6.3.3. Ubicación del trabajo de Pascal-Fermat-Huygens. Hemos visto que no son los trabajos de Pascal-Fermat o el de Huygens los primeros que se refieren al cálculo de probabilidades. Ya antes se habían dado soluciones a algunos problemas particulares e incluso existía ya el trabajo de Girolamo Cardano, el cual contenía un tratamiento más o menos sistemático de problemas con dados. Parece claro que ese trabajo previo que existía había hecho surgir ya la definición clásica de probabilidad, aunque no como una definición general sino solo como un método para resolver determinado tipo de problemas y todavía con una limitación en el sentido de que no se había hecho explícita la necesidad de tener resultados equiprobables para aplicarlo.

El trabajo de Cardano puede considerarse como una síntesis de lo que había antes de Pascal, Fermat y Huygens en lo que se refiere al Cálculo de Probabilidades. Se encuentre en ese trabajo el uso correcto de la definición clásica de probabilidad, una interpretación frecuencial de ésta y la idea de un juego justo; contiene además, aunque usada erróneamente, la idea que está detrás de la regla del producto para el caso de experimentos independientes. Los problemas con dados que trata Cardano son,

³Este problema fue propuesto por Pascal a Fermat.

sin embargo, muy simples; tanto que en ellos la aplicación de la definición clásica es inmediata y no generan, por lo tanto, métodos que puedan convertirse después en una base teórica del Cálculo de Probabilidades. Este hecho justifica tal vez la indiferencia posterior que hubo hacia el trabajo de Cardano.

Echando una vista al trabajo de Cardano y al de Pascal-Fermat-Huygens podemos ver que los problemas que se plantearon los segundos son de una complejidad superior a los que se planteó Cardano. Así, por ejemplo, en el problema de las partidas, aun teniendo a la mano la definición clásica de probabilidad, no resulta evidente la manera de aplicarla; recuérdese que incluso Pascal pensó que resultaba muy complejo resolver el problema por este método y buscó entonces otro. En los problemas con dados, si bien estaba ya resuelto el problema de determinar el número de formas en que pueden caer n dados, no era tampoco evidente el determinar de ahí el número de casos favorables a determinado evento.

La complejidad de los problemas que atacaron Pascal, Fermat y Huygens exigía contar no solo con la definición clásica de probabilidad sino además con métodos o reglas que permitieran simplificar los problemas; la creación implícita de estas reglas o métodos es uno de los méritos del trabajo de Pascal-Fermat-Huygens.

Como hemos visto, las soluciones que dieron a los problemas que se plantearon contienen, implícitamente, prácticamente toda la Teoría del Cálculo de Probabilidades clásico. La abstracción de esta teoría vendría posteriormente pero teniendo como base el trabajo de Pascal-Fermat-Huygens.

Todos los trabajos posteriores sobre el Cálculo de Probabilidades tendrían como base el de Pascal-Fermat-Huygens; Cardano no sería ni siquiera mencionado por los diversos autores. Particularmente importante resultó el trabajo de Huygens, en parte por ser el único que quedó por escrito y en parte por contener algunos conceptos más elaborados que en el trabajo de Pascal-Fermat; aunque hay que tener presente que por ejemplo Fermat es quien hace explícita la necesidad de la equiprobabilidad y quien con mayor claridad expresa la idea de descomposición de un evento en eventos más simples.

También no se puede dejar de señalar que los métodos usados por Pascal y Fermat tienen el mérito de mostrar al Análisis Combinatorio como una herramienta de gran utilidad en los problemas de probabilidad; hecho que más tarde sería sistematizado por Jacques Bernoulli. Siguiendo este camino fructificaría más el Cálculo de Probabilidades pues el método de Huygens, si bien podía generar (y de hecho generó) resultados generales, en casi todos los problemas su uso resultaba sumamente complejo y ésta era su limitación.

Resumiendo, podemos decir que si bien Pascal, Fermat y Huygens no son los primeros en resolver problemas de probabilidad, si es su trabajo de una mayor riqueza y el que influenciaría totalmente todos los trabajos posteriores. La indiferencia total hacia los trabajos anteriores al de Pascal-Fermat-Huygens parece de cualquier manera sumamente injusta pues fue en esos trabajos donde se fue generando la definición clásica de probabilidad. De manera que, si bien todavía a un nivel rudimentario, estos trabajos previos no pueden dejar de considerarse como una etapa en el desarrollo del Cálculo de Probabilidades y, en particular, el trabajo de Cardano como una síntesis de esa etapa.

Referencias

- [1] Bernoulli, J., *L'Art de Conjecturer*, L.G.F. Vastel, G. Le Roy, Caen, 1801. Traducción de *Ars Conjectandi*, Basileae, 1713.
- [2] Cardano, G., *Liber de ludo aleae*, 1564. Publicado en *Opera Imnia*, Vol. 1, 1663. Traducción al inglés en *The book on games on chance*, Holt, Rinehart and Winston, New York, 1961.
- [3] Fermat, P. & Pascal, B., *Correspondance - 1654*, *Oeuvres de Pascal*, t. III, p. 369-430.
- [4] Galileo, G., *Sopra le scoperte dei dadi*, 1620 aprox. Publicado en *Opera Omnia* con el título "Considerazione sopra il giuoco dei dadi", 1718. Traducción al inglés en David, F. N., *Games, goods and gambling - The origins and history of probability and statistical ideas from the earliest times to the Newtonian era*, Griffin, London, 1962.
- [5] Huygens, C., *Du calcul dans les jeux de hasard*, *Oeuvres Complètes de Christiaan Huygens*, Vol. XIV, Martinus Nijhoff, 1920. Traducción de *De Ratiociniis in Aleae Ludo*, 1657.
- [6] Pascal, B., *Traité du triangle arithmétique - 1654*, *Oeuvres Complètes*, Gallimard, 1964.

CAPÍTULO 7

SURGIMIENTO DE LA TEORÍA DE LA PROBABILIDAD MODERNA

Si de un proceso real abstraemos sus aspectos esenciales, dejamos un residuo que debemos considerar aleatorio... Podría pensarse que cuando el “residuo aleatorio”, para una formulación dada de un fenómeno, es tan grande que no se puede despreciar, entonces el único modo posible de proceder sería describir el fenómeno con mayor aproximación... Afortunadamente, nuestras exigencias prácticas son generalmente muy diferentes; solamente necesitamos estimar el efecto total ejercido por los factores aleatorios para un largo intervalo de tiempo o para un gran número de repeticiones del proceso en estudio... Ejemplos en que el efecto de un gran número de factores aleatorios conduce a leyes estadísticas completamente bien definidas, se podrían multiplicar fácilmente. Uno de los más conocidos y al mismo tiempo más fascinante, en vista de la amplitud de sus aplicaciones, es la Teoría Cinética de los Gases, la cual muestra como la influencia conjunta de colisiones aleatorias de moléculas da lugar a leyes precisas que gobiernan la presión de un gas sobre una pared, la difusión de un gas en el seno de otro, etc.

Andrey Nikolaevich Kolmogorov

En el año 1933, Andrey Nikolaevich Kolmogorov publicó un artículo ([63]) en el cual estableció la formulación de la Teoría de la Probabilidad que prevalece hasta nuestros días. En ese artículo, Kolmogorov formuló que el modelo matemático de un fenómeno probabilístico está dado por una terna $(\Omega, \mathfrak{S}, P)$, en donde Ω es un conjunto, \mathfrak{S} una

σ -álgebra de subconjuntos de Ω y P una medida de probabilidad definida sobre \mathfrak{F} . En este capítulo se analizará el proceso que condujo a dicha formulación.

Para comprensión de los elementos que intervienen en el modelo de Kolmogorov, así como para la exposición de la manera en que surge, se requieren las siguientes definiciones:

DEFINICIÓN 7.1 (Álgebra de subconjuntos). *Sea Ω un conjunto. Se dice que una familia \mathcal{A} de subconjuntos de Ω es un álgebra si se satisfacen las siguientes condiciones:*

- (i) $\Omega \in \mathcal{A}$.
- (ii) Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$.
- (iii) Si A_1, \dots, A_n es cualquier familia finita de elementos de \mathcal{A} , entonces $\bigcup_{k=1}^n A_k \in \mathcal{A}$.

DEFINICIÓN 7.2 (σ -álgebra de subconjuntos). *Sea Ω un conjunto. Se dice que una familia \mathcal{A} de subconjuntos de Ω es un σ -álgebra si es un álgebra y dada cualquier familia infinita numerable de elementos de \mathcal{A} , A_1, A_2, \dots , entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.*

DEFINICIÓN 7.3 (Conjuntos ajenos por parejas). *Dada una familia de conjuntos $(A_\gamma)_{\gamma \in \Gamma}$, se dice que éstos son ajenos por parejas si $A_\alpha \cap A_\beta = \emptyset$ para $\alpha, \beta \in \Gamma$ y $\alpha \neq \beta$.*

Obsérvese que como un álgebra es cerrada bajo uniones y complementos, también lo es bajo intersecciones y diferencias, de manera que para demostrar que un álgebra \mathcal{A} es una σ -álgebra basta con probar que si A_1, A_2, \dots es cualquier colección numerable de elementos de \mathcal{A} , ajenos por parejas, entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$.

DEFINICIÓN 7.4 (Propiedad de la aditividad finita). *Sea Ω un conjunto y \mathcal{A} un álgebra de subconjuntos de Ω . Se dice que una función no negativa P , definida sobre \mathcal{A} , es finitamente aditiva (o que tiene la propiedad de la aditividad finita) si dada cualquier familia finita, A_1, \dots, A_n , de elementos de \mathcal{A} , ajenos por parejas, entonces:*

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

DEFINICIÓN 7.5 (Propiedad de la aditividad numerable). *Sea Ω un conjunto y \mathcal{A} un álgebra de subconjuntos de Ω . Se dice que una función no negativa P , definida sobre \mathcal{A} , es σ -aditiva (o que tiene la propiedad de la aditividad numerable) si es finitamente aditiva y dada cualquier familia infinita numerable, A_1, A_2, \dots , de elementos de \mathcal{A} , ajenos por parejas, tales que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces:*

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

DEFINICIÓN 7.6 (Medidas de probabilidad). *Sea Ω un conjunto y \mathcal{A} una σ -álgebra de subconjuntos de Ω . Se dice que una función no negativa P , definida sobre \mathcal{A} , es una medida de probabilidad si es σ -aditiva y $P(\Omega) = 1$.*

En el modelo de Kolmogorov se identifica a una función de probabilidad con una medida en el sentido del Análisis Matemático, concepto que surgió y se desarrolló en los primeros 30 años del siglo XX. Sin embargo, tal identificación no surgió de manera automática como algo general aplicable a cualquier experimento aleatorio sino que requirió de un proceso, que llevó varios años, en el cual se mostró que es una alternativa adecuada. En el centro de este proceso se encuentra el planteamiento de problemas en donde se trata de calcular probabilidades de eventos cuya ocurrencia o no ocurrencia depende de una infinidad de observaciones y la aceptación de la aditividad numerable como una propiedad general de la función de probabilidad, la cual permite atacar ese tipo de problemas.

Con el nacimiento de la Teoría de la Medida de Borel-, durante 1898-1902, se comenzaron a identificar funciones de probabilidad con medidas pero únicamente en cierto tipo de modelos probabilísticos, aquellos que caían dentro de un esquema geométrico. Incluso cuando ya se había iniciado el desarrollo de una teoría general de la medida, alrededor del año 1915, no se dio mecánicamente la identificación de una función de probabilidad cualquiera con una medida, siendo la no aceptación de la σ -aditividad uno de los impedimentos. No fue sino hasta la publicación del trabajo de Kolmogorov cuando la σ -aditividad comenzó a ser ampliamente aceptada, siendo varios los factores que influyeron para tal aceptación. Por un lado, la Teoría de la Medida se había desarrollado con suficiente generalidad, permitiendo así definir medidas en cualquier conjunto. Por otro lado, se había ampliado el marco de aplicabilidad de la σ -aditividad hasta abarcar problemas básicos de la Teoría de la Probabilidad como son los teoremas límite para sucesiones de variables aleatorias independientes. Pero el elemento central consistió en el hecho de que se mostró que la σ -aditividad permite la construcción de modelos matemáticos en problemas que involucran a una infinidad de variables aleatorias, siendo los trabajos de Hugo Dyonizy Steinhaus ([96]) y Norbert Wiener ([102], [103], [104], [105], [106], [107], [108]) los que abrieron el camino en esta dirección, hasta llegar al resultado general de Kolmogorov, quien demostró que, aceptando la σ -aditividad en el caso de un número finito de variables aleatorias, siempre es posible extenderla al caso de una familia arbitraria, mostrando así la consistencia del tratamiento matemático de los fenómenos probabilísticos asumiendo que la probabilidad es una medida.

Un aspecto importante a resaltar consiste en que, al analizar el proceso que conduce al modelo de Kolmogorov, puede verse que **la aditividad numerable surge como una herramienta matemática, la cual permite extender, de manera única, una función de probabilidad definida para una cierta familia de eventos a una familia mucho más amplia.**

Por otra parte, debe de observarse que, en general, la σ -aditividad no es una consecuencia de la aditividad finita. Consideremos, por ejemplo, el álgebra \mathcal{A} formada por los subconjuntos de los números naturales que son finitos o de complemento finito y definamos la función $P : \mathcal{A} \mapsto [0, 1]$ por $P(A) = 0$ si A es finito y $P(A) = 1$ si A^c es finito. Tal función es finitamente aditiva pero no σ -aditiva. Más aún, se puede mostrar ([90]) que P puede extenderse (no de manera única) a una función finitamente aditiva definida sobre la familia de todos los subconjuntos de los números naturales. Tal extensión, la cual está definida sobre una σ -álgebra, resulta entonces ser finitamente aditiva, pero no σ -aditiva.

Debe mencionarse también que no es Kolmogorov el primero en plantear un modelo matemático en donde es aceptada la σ -aditividad como una propiedad general de cualquier función de probabilidad. El planteamiento más completo en este sentido, previo al trabajo de Kolmogorov, se debe a Paul Pierre Lévy, quien en su libro “Calcul des Probabilités”, publicado en 1925 ([74], [75]), da incluso un método para definir funciones σ -aditivas en espacios de dimensión infinita. Sin embargo, el método de Lévy no era lo suficientemente general como para abarcar cualquier fenómeno probabilístico.

7.1. El Cálculo de Probabilidades clásico

A principios del siglo XX la Teoría de la Probabilidad gozaba ya de una gran popularidad. Por un lado, los trabajos de Pierre Simon Laplace ([67], [68], [69], [70], [71]), eran ya ampliamente conocidos en el medio científico. En particular, en su “Teoría Analítica de las Probabilidades”, Laplace logró sistematizar los métodos para calcular probabilidades. Por otro lado, la escuela de San Petersburgo, formada, entre otros, por Pafnuty Lvovich Chebyshev, Andrei Andreyevich Markov y Aleksandr Mikhailovich Lyapunov, había hecho aportaciones claves ([34], [35], [36], [80], [81], [82], [83], [84], [78] y [79]), las cuales conducirían a la forma general de la Ley Débil de los Grandes Números y del Teorema del Límite Central. Finalmente, además de las aplicaciones al estudio de datos estadísticos, la Teoría de la Probabilidad estaba siendo aplicada en la solución de problemas importantes de la Física, como son los

referentes a la Mecánica Estadística, con los trabajos de A. Krönig, Rudolf Julius Emmanuel Clausius, James Clerk Maxwell, Ludwig Boltzmann y Josiah Willard Gibbs ([64], [31], [32], [33], [85], [86], [5], [6], [7], [8], [9], [10], [11], [52])¹.

Sin embargo, los fundamentos matemáticos de la Teoría de la Probabilidad no eran satisfactorios. De hecho, la Probabilidad no era considerada como parte de la Matemática. Sus conceptos y métodos eran específicos para las aplicaciones y no formaban parte de una estructura abstracta general. La misma definición de probabilidad, la cual estaba basada en el concepto de equiprobabilidad, resultaba insatisfactoria pues no en todos los fenómenos aleatorios resulta evidente qué resultados pueden considerarse como equiprobables.

Como muestra de la visión que se tenía sobre la probabilidad a principios de este siglo basta citar el artículo de David Hilbert ([58]) presentado en el décimo Congreso Internacional de Matemáticas, realizado en el año 1900, en donde afirmó: “pienso que en cualquier lugar en donde se presenten ideas matemáticas, sea en Filosofía, sea en Geometría, sea en Física, se plantea el problema de la discusión de los principios fundamentales, base de esas ideas, y del establecimiento de un sistema simple y completo de axiomas”; y más adelante continúa: “Las investigaciones sobre los principios fundamentales de la geometría nos conducen a plantear este problema: Tratar con base en ese modelo las ramas de la Física donde las Matemáticas juegan actualmente un papel preponderante; esas ramas de la ciencia son, antes que cualesquiera otras, el Cálculo de Probabilidades y la Mecánica”.

Una buena referencia para conocer el estado de la Teoría de la Probabilidad a principios del siglo XX es el libro de Jules Henri Poincaré ([88]).

Lo primero que resalta en el libro de Poincaré es su primera frase: “No se puede dar una definición satisfactoria de la probabilidad”. En seguida enuncia la ahora llamada definición clásica de probabilidad: “la probabilidad de un evento es el cociente de los casos favorables a un evento y el número total de casos posibles”, aclarando mediante algunos ejemplos que se debe agregar a dicha definición la condición de que todos los casos sean igualmente probables. Comenta entonces que “la definición completa de la probabilidad es una especie de petición de principio: ¿cómo reconocer que todos los casos son igualmente probables? Aquí, una definición matemática no es posible; deberemos, en cada aplicación, hacer convenciones, decir que consideramos tal y tal caso como igualmente probables. Esas convenciones no son completamente arbitrarias, pero escapan al espíritu del matemático que no tendrá más que examinarlas, una vez que son admitidas. Así, todo problema de probabilidad ofrece dos periodos de

¹Una muy buena exposición analítica sobre la utilización de la Teoría de la Probabilidad en la Física Estadística, así como en la Mecánica Cuántica, puede verse en [101].

estudio: el primero, metafísico por así decirlo, el cual legitima tal o cual convención; el segundo, matemático, que aplica a esas convenciones las reglas del cálculo”.

Para Poincaré, el Cálculo de Probabilidades estaba basado en dos teoremas: el de probabilidades totales y el de probabilidades compuestas, los cuales no son otra cosa que las reglas de la suma y del producto para dos eventos A y B :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

Poincaré hacía la demostración de estos teoremas aplicando la definición clásica de probabilidad, asumiendo, en particular, que el total de casos posibles es finito.

Utilizando estas reglas y la definición de probabilidad de un evento, Poincaré resolvió una buena cantidad de problemas, algunos del tipo de los que uno encuentra actualmente en los textos introductorios a la Teoría de la Probabilidad, otros más complicados, pero cuyas soluciones están basadas en los mismos métodos.

Consideraba Poincaré que hay tres tipos de problemas en la Teoría de la Probabilidad. Dentro de la primera categoría están aquellos en los cuales el número de casos posibles es finito y no sobrepasa ciertos límites, en cuyo caso, decía, se trata, en general, de juegos de azar y de problemas simples de Cálculo Combinatorio. Dentro de la segunda categoría están aquellos en los cuales el número de casos posibles es finito pero se hace muy grande, en cuyo caso se tiene únicamente una expresión aproximada de la probabilidad mediante la Ley de los Grandes Números, el Teorema de Bernoulli, etc. Finalmente, dentro de la tercera categoría están aquellos en los cuales el número de casos posibles es infinito, en cuyo caso las probabilidades relativas a una cantidad x están determinadas por una función φ , de una o varias variables, de tal manera que la probabilidad de que x pertenezca a una región A está dada por $\int_A \varphi(x)dx$.

Siguiendo a Poincaré, dicha función φ deberá darse al inicio del problema mediante una convención especial y será, en general, una función continua. Los problemas que caen dentro de esta categoría los llamaba Poincaré de probabilidades continuas.

Como puede verse, el tratamiento de los problemas que consideraba Poincaré en los cuales el número de posibles resultados es infinito obedece a un esquema geométrico, considerándose en esta categoría problemas como el de la aguja de Buffon y la paradoja de Bertrand.

Debe mencionarse sin embargo que implícitamente estaba presente una clase de problemas en los cuales el número de casos posibles es infinito sin ser de probabilidades continuas. Por ejemplo Poincaré consideró el problema de los 3 jugadores:

PROBLEMA 1. *Tres jugadores, P, Q y R, juegan partidas por parejas en cada una de las cuales la probabilidad que cada jugador tiene de ganar es $\frac{1}{2}$; quien gane una partida juega con el otro jugador hasta que uno de los jugadores gane dos partidas consecutivas, ganando entonces el juego. Suponiendo que comienzan jugando P contra Q, encontrar las probabilidades que cada uno tiene de ganar el juego.*

Poincaré resolvió este problema aplicando las reglas de la suma y del producto, las cuales había demostrado para el caso en que el número de posibles resultados es finito. El razonamiento es el siguiente:

Sean A, B, C los eventos consistentes en que P, Q y R ganen el juego respectivamente. Sean además x, y, z las probabilidades condicionales, dado que P ganó la primera partida, de que P, Q y R ganen el juego respectivamente. Entonces, aplicando la regla de la probabilidad total, se obtiene:

$$x = \frac{1}{2} + \frac{1}{2}y, \quad y = \frac{1}{2}z, \quad z = \frac{1}{2}x$$

Por lo tanto: $x = \frac{4}{7}, \quad y = \frac{1}{7}, \quad z = \frac{2}{7}.$

Finalmente, aplicando nuevamente la regla de la probabilidad total, se tiene:

$$P(A) = P(B) = \frac{1}{2}x + \frac{1}{2}y = \frac{5}{14}$$

$$P(C) = \frac{1}{2}z + \frac{1}{2}z = \frac{4}{14}$$

También consideró Poincaré el problema de la ruina del jugador:

PROBLEMA 2. *Dos personas A y B juegan partidas en cada una de las cuales la primera tiene una probabilidad p de ganarla y una probabilidad $1 - p$ de perderla. Inicialmente A y B tienen m y n fichas respectivamente y después de cada partida el perdedor entrega una ficha al vencedor. Si el juego se termina en el momento en que uno de los dos jugadores quede arruinado, calcular la probabilidad de que éste sea A y la probabilidad de que sea B el que lo haga.*

Nuevamente aplicando las reglas de la suma y del producto, Poincaré resolvió este problema de la siguiente manera:

Sea $\varphi(k)$ la probabilidad de que B termine arruinado comenzando con k fichas, entonces aplicando la regla de la probabilidad total, se obtiene:

$$\varphi(k) = p\varphi(k - 1) + (1 - p)\varphi(k + 1)$$

Esta es una ecuación en diferencias finitas, con las condiciones iniciales $\varphi(0) = 1$, $\varphi(s) = 0$, en donde $s = m + n$.

Resolviendo esta ecuación, se encuentra el valor siguiente para φ :

$$\varphi(k) = \begin{cases} \frac{s-k}{2} & \text{si } p = \frac{1}{2} \\ \frac{\beta^k - \beta^s}{1 - \beta^s} & \text{en otro caso} \end{cases}$$

en donde $\beta = \frac{p}{1-p}$.

La utilización de reglas demostradas para el caso en que el número de posibles resultados es finito a problemas en donde el número de posibles resultados es infinito no es algo novedoso en el libro de Poincaré, así se hacía desde muy al principio del Cálculo de Probabilidades, siendo Christiaan Huygens el primero en hacer un razonamiento de este tipo al plantear y resolver el siguiente problema ([59]):

PROBLEMA 3. *Dos jugadores, P y Q, juegan a lanzar alternadamente un par de dados. El juego comienza lanzando P el par de dados, con la condición de que si obtiene una suma igual a 6 gana el juego, en caso contrario el juego continúa lanzando Q el par de dados, con la condición de que si obtiene una suma igual a 7 gana el juego, en caso contrario el juego continúa lanzando P el par de dados bajo las condiciones iniciales. ¿Cuáles son las respectivas probabilidades que cada jugador tiene de ganar el juego?*

El razonamiento de Huygens para resolver este problema era esencialmente² el siguiente:

Definamos los siguientes eventos:

A: El juego es ganado por P.

B: El juego es ganado por Q.

*A*₁: P obtiene una suma igual a 6 en su primer lanzamiento.

E: P fracasa en su primer lanzamiento y Q obtiene una suma igual a 7 en su primer lanzamiento.

F: Tanto P como Q fracasan en su primer intento.

Entonces:

$$P(A) = P(A | A_1)P(A_1) + P(A | E)P(E) + P(A | F)P(F)$$

$$= P(A_1) + P(A)P(F) = \frac{5}{36} + \frac{31}{36} \frac{30}{36} P(A)$$

$$P(B) = P(B | A_1)P(A_1) + P(B | E)P(E) + P(B | F)P(F)$$

²Huygens no consideraba probabilidades sino esperanzas.

$$= P(E) + P(B)P(F) = \frac{31}{36} \frac{6}{36} + \frac{31}{36} \frac{30}{36} P(B)$$

Así que, $P(A) = \frac{30}{61}$ y $P(B) = \frac{31}{61}$.

Más tarde, Jacques Bernoulli ([1]) resolvería este problema estableciendo una progresión geométrica para la Esperanza que cada jugador tiene sobre lo que se gana en el juego. En términos de probabilidades, el resultado de Bernoulli puede escribirse como sigue:

En cada lanzamiento del par de dados, diremos que hay éxito si el jugador que los está lanzando gana el juego en ese lanzamiento. Un posible resultado para este experimento aleatorio puede ser representado entonces por una sucesión finita, (F, \dots, F, S) , compuesta de fracasos consecutivos seguidos de un éxito, o por una sucesión infinita, (F, F, \dots) , compuesta exclusivamente de fracasos. Denotemos por ω_n al resultado (F, \dots, F, S) , compuesto de $n - 1$ fracasos seguidos de un éxito y por $p(\omega_n)$ a la probabilidad de ocurrencia de ω_n . Se tiene entonces:

$$p(\omega_{2k}) = (1 - p_1)^k (1 - p_2)^{k-1} p_2 \quad \text{si } n = 2k \text{ para algún número natural } k$$

$$p(\omega_{2k-1}) = (1 - p_1)^{k-1} (1 - p_2)^{k-1} p_1 \quad \text{si } n = 2k - 1 \text{ para algún número natural } k$$

en donde p_1, p_2 son las probabilidades de obtener 6 y 7, respectivamente, al lanzar un par de dados.

Ahora bien, considerando que P gana cuando ocurre ω_n para algún n impar, mientras que Q gana cuando ocurre ω_n para algún n par, se tiene:

$$P(A) = \sum_{k=1}^{\infty} p(\omega_{2k-1}) = \sum_{k=1}^{\infty} \left(\frac{31}{36}\right)^{k-1} \left(\frac{30}{36}\right)^{k-1} \frac{5}{36} = \frac{30}{61}$$

$$P(B) = \sum_{k=1}^{\infty} p(\omega_{2k}) = \sum_{k=1}^{\infty} \left(\frac{31}{36}\right)^k \left(\frac{30}{36}\right)^{k-1} \frac{6}{36} = \frac{31}{61}$$

Bernoulli estaba estableciendo entonces que la probabilidad de que un jugador gane el juego es igual a la suma de las probabilidades de que gane en cada uno de los posibles turnos que tiene, los cuales son una infinidad. En otras palabras, está implícita en el resultado la propiedad de la aditividad numerable de la función de probabilidad.

Bernoulli hacía ver la limitación que tiene el método de Huygens, el cual puede aplicarse únicamente cuando un juego consiste de una sucesión continua de tiradas y después de cierto número de ellas reaparece la misma situación que se tenía al comenzar el juego, en cuyo caso, dice Bernoulli, es recomendable usarlo. Hacía entonces hincapié en que su método puede aplicarse aún en el caso en que cada vez se encuentren “suertes” nuevas hasta el infinito. Para ejemplificar, planteó los siguientes problemas:

PROBLEMA 4. *A y B juegan con dos dados y el primero que obtenga la suma 7 ganará el juego. Calcular la probabilidad que cada uno tiene en el juego con cada una de las siguientes condiciones en el orden de las tiradas:*

A tira una vez, B una vez, A dos veces, B una vez, A tres veces, B una vez, etc.

A tira una vez, B una vez, A una vez, B dos veces, A una vez, B tres veces, etc.

A tira una vez, B una vez, A dos veces, B dos veces, A tres veces, B tres veces, etc.

A tira una vez, B dos veces, A tres veces, B cuatro veces, A cinco veces, B seis veces, etc.

Bernoulli resolvió estos problemas aplicando su método, con el cual obtenía una serie para la probabilidad que cada jugador tiene de ganar el juego en cada una de las situaciones planteadas.

El método de Bernoulli fue retomado un poco más adelante, en el año 1718, por Abraham de Moivre en su libro ([44]), sin embargo, aunque aparentemente era conocido, no se utilizó durante el resto del siglo XVIII y todo el XIX, de manera que la propiedad de la aditividad numerable de la función de probabilidad quedó relegada en la sistematización de Laplace, la cual perduró hasta principios del siglo XX.

El poco interés que atrajo el método de Bernoulli puede no haber sido circunstancial, sino que parece obedecer a la concepción de la probabilidad que está implícita en su formulación clásica, la cual está basada en la equiprobabilidad de los diferentes resultados de un experimento aleatorio, cuyo número debe entonces ser finito. De manera que los problemas en donde se realizan repeticiones indefinidas de experimentos aleatorios únicamente pueden tratarse mediante aproximaciones a través de sus correspondientes límites. Los problemas considerados arriba, en donde el número de casos posibles es infinito sin ser de probabilidades continuas, caen dentro de esta categoría de problemas, que no rebasan el marco clásico. En efecto, la solución de Bernoulli al problema planteado por Huygens, por ejemplo, puede plantearse como una distribución límite considerando las probabilidades que cada jugador tiene de ganar en los primeros $2n$ lanzamientos y haciendo tender luego n a ∞ . Si llamamos $P_n(A)$ y $P_n(B)$ a estas probabilidades se obtiene, utilizando la notación mencionada más arriba:

$$P_n(A) = \sum_{k=1}^n p(\omega_{2k-1}) = \sum_{k=1}^n \left(\frac{31}{36}\right)^{k-1} \left(\frac{30}{36}\right)^{k-1} \frac{5}{36} = \frac{30}{61} \left[1 - \left(\frac{31}{36}\right)^n \left(\frac{30}{36}\right)^n\right]$$

$$P_n(B) = \sum_{k=1}^n p(\omega_{2k}) = \sum_{k=1}^n \left(\frac{31}{36}\right)^k \left(\frac{30}{36}\right)^{k-1} \frac{6}{36} = \frac{31}{61} \left[1 - \left(\frac{31}{36}\right)^n \left(\frac{30}{36}\right)^n\right]$$

Veremos más adelante que efectivamente, incluso todavía en los 20's del siglo pasado se daba esta interpretación a la solución de Bernoulli.

El marco clásico fue rebasado en el momento en que se plantearon y resolvieron problemas relativos a eventos cuya ocurrencia no sólo dependiera de una infinidad de ensayos sino que, además, sus probabilidades no pudieran ser vistas como una distribución límite cuando el número de ensayos tiende a ∞ . Problemas de este tipo son los que planteó Félix Édouard Justin Émile Borel en un artículo publicado en el año 1909, abriendo así el camino hacia el abandono de la formulación clásica de la Teoría de la Probabilidad.

7.2. Las probabilidades numerables de Émile Borel

Borel puso el dedo en la llaga sobre el problema de la aditividad numerable de la función de probabilidad. En su artículo ‘ ‘Les probabilités dénombrables et leurs applications arithmétiques’’, publicado en el año 1909 ([16]), decía Borel:

“Se distinguen generalmente, en los problemas de probabilidad, dos categorías principales, dependiendo de que el número de casos posibles sea finito o infinito: la primera categoría constituye lo que se llama las *probabilidades discontinuas*, o probabilidades en el dominio del discontinuo, mientras que la segunda categoría comprende las *probabilidades continuas o probabilidades geométricas*. Tal clasificación aparece como incompleta cuando se consideran los resultados de la Teoría de Conjuntos; entre la potencia de los conjuntos finitos y la potencia del continuo se encuentra la potencia de los conjuntos numerables; me propongo mostrar brevemente el interés respecto a las cuestiones de probabilidad en cuyo enunciado intervienen tales conjuntos; las llamaré, para abreviar, *probabilidades numerables*”.

En seguida clasificaba Borel los problemas de probabilidades numerables en tres categorías: los de primera categoría son aquellos en los cuales el número de resultados de cada ensayo es finito, pero el número de ensayos es infinito numerable. En los problemas de segunda categoría, el número de resultados en cada ensayo constituye una infinidad numerable, pero el número de ensayos es finito. Finalmente, en los problemas de tercera categoría, tanto el número de resultados en cada ensayo como el número de ensayos constituyen una infinidad numerable.

Las ideas esenciales que aportó Borel en su artículo están contenidas en el análisis que hace de los problemas de primera categoría. Por tal motivo, basta con examinar este caso. De igual forma, es suficiente con considerar el caso en que cada ensayo admite únicamente dos posibles resultados, los cuales serán llamados éxito y fracaso. Finalmente, aunque Borel no lo hacía explícito, se asume que los ensayos son independientes unos de otros.

Consideremos entonces una sucesión infinita numerable de ensayos independientes y denotemos por p_n a la probabilidad de éxito en el ensayo n . El problema principal

que atacó Borel en este caso consiste en determinar la probabilidad de que se obtenga éxito una infinidad de veces. Para atacar este problema, Borel se planteó primero el problema de determinar, para cada $k \in \{0, 1, \dots\}$, la probabilidad de que se obtenga éxito exactamente k veces en la sucesión infinita de ensayos. Definamos entonces los eventos:

A_k : Se obtienen exactamente k éxitos en la infinidad de ensayos.

Como puede verse, la ocurrencia o no ocurrencia de los eventos A_k depende de la infinidad de ensayos, de manera que sus probabilidades no pueden obtenerse de manera directa con las reglas que se aplican en el caso finito.

En general, para encontrar la probabilidad de un evento, cuya ocurrencia o no ocurrencia depende de la infinidad de ensayos, Borel resolvía el mismo problema pero asumiendo que el número de ensayos es n , después de lo cual hacía tender n a infinito. Para seguir este método, definamos, para cada $k \in \{0, 1, \dots\}$ y $n \in \mathbb{N}$, el evento:

A_k^n : Se producen exactamente k éxitos en los primeros n ensayos.

Para simplificar el análisis asumiremos que $0 \leq p_n < 1$ para cualquier n . Por otra parte, antes de entrar propiamente a los cálculos de Borel, conviene analizar las dos situaciones bajo las cuales se realizan.

1er. caso: la serie $\sum_{n=1}^{\infty} p_n$ es convergente.

En este caso, por el teorema del valor medio, se tiene, para cada $x \in [0, 1)$:

$$\ln(1 - x) = -\frac{x}{1 - \theta x}, \text{ en donde } \theta \in (0, 1)$$

En particular, para cada $n \in \mathbb{N}$, se tiene:

$$\ln(1 - p_n) = -\frac{p_n}{1 - \theta_n p_n}, \text{ en donde } \theta_n \in (0, 1)$$

Pero, como $\lim_{n \rightarrow \infty} p_n = 0$, existe $N \in \mathbb{N}$ tal que $p_n < \frac{1}{2}$ para cualquier $n \geq N$. Así que, para cualquier $n \geq N$, $\frac{1}{1 - \theta_n p_n} < 2$. Por lo tanto, $\ln(1 - p_n) > -2p_n$, es decir, $1 - p_n > e^{-2p_n}$. De manera que, si definimos $a_n = \prod_{j=1}^n (1 - p_j)$, entonces $a_n > A_N e^{-2 \sum_{j=N}^n p_j}$, en donde $A_N = \prod_{j=1}^{N-1} (1 - p_j) > 0$. Así que:

$$\lim_{n \rightarrow \infty} a_n \geq A_N \lim_{n \rightarrow \infty} e^{-2 \sum_{j=N}^n p_j} = A_N e^{-2 \sum_{j=N}^{\infty} p_j} > 0$$

Además, como también se tiene $\frac{1}{1 - p_n} < 2$ para cualquier $n \geq N$, entonces $\sum_{n=N}^{\infty} \frac{p_n}{1 - p_n} < 2 \sum_{n=N}^{\infty} p_n$, así que también la serie $\sum_{n=1}^{\infty} \frac{p_n}{1 - p_n}$ es convergente.

Por otra parte, $1 + \frac{p_n}{1 - p_n} = \frac{1}{1 - p_n}$ para cualquier n , de manera que:

$$\lim_{n \rightsquigarrow \infty} \prod_{j=1}^n \left(1 + \frac{p_j}{1-p_j}\right) = \lim_{n \rightsquigarrow \infty} \prod_{j=1}^n \frac{1}{1-p_j} = \frac{1}{\lim_{n \rightsquigarrow \infty} a_n}$$

2o. caso: la serie $\sum_{n=1}^{\infty} p_n$ es divergente.

Nuevamente, por el teorema del valor medio, se tiene, para cada $x \in [0, 1)$:

$$\ln(1-x) = -\frac{x}{1-\theta x}, \text{ en donde } \theta \in (0, 1)$$

De manera que, para cualquier $x \in [0, 1)$, $\ln(1-x) \leq -x$, es decir, $1-x \leq e^{-x}$.

En particular, $\prod_{j=1}^n (1-p_j) < e^{-\sum_{j=1}^n p_j}$, así que $\lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1-p_j) = 0$.

Además, como también se tiene $\sum_{n=1}^{\infty} \frac{p_n}{1-p_n} > \sum_{n=1}^{\infty} p_n$, entonces también la serie $\sum_{n=1}^{\infty} \frac{p_n}{1-p_n}$ es divergente.

En este caso, para simplificar aún más el análisis asumiremos que existen números reales r y s tales que $0 < r < p_n < s < 1$ para cualquier n . Entonces:

$$\frac{nr}{1-r}(1-s)^n < \prod_{j=1}^n (1-p_j) \sum_{j=1}^n \frac{p_j}{1-p_j} < \frac{ns}{1-s}(1-r)^n$$

$$\text{Así que } \lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1-p_j) \sum_{j=1}^n \frac{p_j}{1-p_j} = 0.$$

En resumen, se tiene el siguiente resultado:

LEMA 7.7. *Sea $\{p_n\}$ una sucesión de números reales en el intervalo $[0, 1)$, supongamos que existen números reales r y s tales que $0 < r < p_n < s < 1$ para cualquier $n \in \mathbb{N}$ y, para cada $n \in \mathbb{N}$, definamos $u_n = \frac{p_n}{1-p_n}$.*

- (i) *Si la serie $\sum_{n=1}^{\infty} p_n$ es convergente, entonces $a = \lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1-p_j) > 0$, la serie $\sum_{n=1}^{\infty} u_n$ es convergente y $\lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1+u_j) = \frac{1}{a}$.*
- (ii) *Si la serie $\sum_{n=1}^{\infty} p_n$ es divergente, entonces $\lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1-p_j) = 0$, la serie $\sum_{n=1}^{\infty} u_n$ es divergente y $\lim_{n \rightsquigarrow \infty} \prod_{j=1}^n (1-p_j) \sum_{j=1}^n u_j = 0$.*

Siguiendo a Borel, pasemos ahora a calcular las probabilidades de los eventos A_k .

Para calcular $P(A_0)$ se requiere calcular primero $P(A_0^n)$ para cualquier n . Esto no presenta ninguna dificultad, como el mismo Borel lo mencionaba: “los principios clásicos permiten definir y calcular la probabilidad de que el caso favorable no se presente en los primeros n ensayos”. En efecto, $P(A_0^n)$ se obtiene inmediatamente aplicando el principio de las probabilidades compuestas, obteniéndose $P(A_0^n) = \prod_{j=1}^n (1-p_j)$.

En el caso en que la serie $\sum_{n=1}^{\infty} p_n$ es convergente, el producto $\prod_{j=1}^n (1-p_j)$ converge a un valor $a > 0$ cuando $n \rightsquigarrow \infty$. Dice Borel que entonces se puede escribir $P(A_0) = a$, presentando la siguiente justificación: “se constata que, cuando n crece, no solamente

esta probabilidad (la de $P(A_0^n)$) varía muy poco de una manera absoluta, sino que varía también muy poco de una manera relativa, es decir que sus variaciones son una fracción muy pequeña de su valor. Se puede entonces, habiendo asignado un valor a la precisión relativa que se dese alcanzar, estar seguro de que esta precisión se alcanza efectivamente al cabo de cierto número de ensayos, quizá muy grande, pero asignable: el límite que hemos efectuado no presenta entonces ninguna dificultad y está completamente justificado”.

En el caso en que la serie $\sum_{n=1}^{\infty} p_n$ es divergente, el producto $\prod_{j=1}^n (1 - p_j)$ converge a 0 cuando $n \rightsquigarrow \infty$. Decía Borel que también en este caso se puede escribir $P(A_0) = 0$, pero haciendo la siguiente aclaración: “hay, en efecto, una verdadera discontinuidad entre una probabilidad infinitamente pequeña, es decir una probabilidad variable que tiende hacia cero, y una probabilidad igual a 0. En efecto, por pequeña que sea la probabilidad del caso favorable, éste es posible; mientras que es imposible si la probabilidad es nula. Tales son al menos los resultados clásicos en la teoría de las probabilidades discontinuas; se sabe que no es lo mismo en la teoría de las probabilidades continuas: la probabilidad de que un número tomado al azar sea racional es nula; eso no quiere decir que no haya números racionales. Así será también en la teoría de las probabilidades numerables: probabilidad nula no deberá ser considerada como el equivalente de imposibilidad. Estando esto bien entendido, ya no hay inconveniente para decir que, en el caso divergente, $P(A_0)$ es nula; pero no deberá perderse de vista que ese lenguaje no significa otra cosa que esto: la probabilidad para que el caso favorable no se produzca tiende hacia cero cuando el número de ensayos aumenta indefinidamente”.

Este comentario de Borel deja ver que para él la probabilidad $P(A_0)$ en el caso divergente no era una verdadera probabilidad, sino únicamente un límite.

Para calcular $P(A_1)$ consideraba Borel los eventos:

C_j : Se produce éxito exclusivamente en el ensayo j .

Entonces:

$$P(C_j) = (1 - p_1) \cdots (1 - p_{j-1}) p_j (1 - p_{j+1}) \cdots = u_j \prod_{j=1}^{\infty} (1 - p_j)$$

De manera que, si la serie $\sum_{n=1}^{\infty} p_n$ es convergente, $P(C_j) = a u_j$, mientras que si la serie $\sum_{n=1}^{\infty} p_n$ es divergente, entonces $P(C_j) = 0$.

En el caso convergente, decía Borel que mediante una justificación similar a la que hace para escribir $P(A_0) = a$, el principio de las probabilidades totales permite escribir:

$$P(A_1) = \sum_{j=1}^{\infty} P(C_j) = a \sum_{j=1}^{\infty} u_j$$

Obsérvese que la expresión $P(A_1) = \sum_{j=1}^{\infty} P(C_j)$ expresa la propiedad de σ -aditividad. Sin embargo, Borel no estaba asumiendo aquí su validez, sino que obtenía la expresión $P(A_1) = a \sum_{j=1}^{\infty} u_j$ con otro método. La justificación la desarrolló Borel únicamente para el caso divergente, pero se puede aplicar en los dos casos. En efecto, definamos los eventos:

C_j^n : En los primeros n ensayos, se produce éxito exclusivamente en el ensayo j .

Entonces:

$$\begin{aligned} P(A_1^n) &= \sum_{j=1}^n P(C_j^n) = \sum_{j=1}^n (1 - p_1) \cdots (1 - p_{j-1}) p_j (1 - p_{j+1}) \cdots (1 - p_n) \\ &= \sum_{j=1}^n u_j (1 - p_1) \cdots (1 - p_n) = \prod_{j=1}^n (1 - p_j) \sum_{j=1}^n u_j \end{aligned}$$

Así que, como $\lim_{n \rightarrow \infty} P(A_1^n) = \lim_{n \rightarrow \infty} \prod_{j=1}^n (1 - p_j) \sum_{j=1}^n u_j = a \sum_{j=1}^{\infty} u_j > 0$, entonces, siguiendo a Borel, no existe dificultad para escribir:

$$P(A_1) = \lim_{n \rightarrow \infty} P(A_1^n) = a \sum_{j=1}^{\infty} u_j$$

Consideraba Borel que la extensión de la última fórmula al caso divergente requiere precaución pues el límite $\lim_{n \rightarrow \infty} \prod_{j=1}^n (1 - p_j) \sum_{j=1}^n u_j$ queda como un producto $0 \cdot \infty$, el cual está indeterminado. Decía también que se puede analizar el problema considerando que $P(C_j) = 0$ para cualquier j y entonces, tomando $P(A_1) = \sum_{j=1}^{\infty} P(C_j)$, se concluiría que $P(A_1) = 0$. Sin embargo, recordando que probabilidad nula no significa imposibilidad, agregaba que no se puede concluir sin precaución que $P(A_1) = 0$.

Obsérvese que aquí Borel estaba diciendo, implícitamente, que la propiedad de σ -aditividad podría no ser válida.

Prefería entonces calcular $P(A_1^n)$ como hicimos antes en el caso convergente, obteniéndose:

$$P(A_1^n) = \prod_{j=1}^n (1 - p_j) \sum_{j=1}^n u_j$$

De manera que:

$$P(A_1) = \lim_{n \rightarrow \infty} P(A_1^n) = \lim_{n \rightarrow \infty} \prod_{j=1}^n (1 - p_j) \sum_{j=1}^n u_j = 0$$

De manera similar se encuentra que, en el caso convergente, $P(A_2) = a \sum_{\{i,j \in \mathbb{N}: i \neq j\}} u_i u_j$, mientras que en el caso divergente, $P(A_2) = 0$.

En general se obtiene, para cualquier $k \in \mathbb{N}$, $P(A_k) = a \sum_{\{(j_1, \dots, j_k) \in C(k)\}} u_{j_1} \cdots u_{j_k}$ en el caso convergente y $P(A_k) = 0$ en el caso divergente, en donde $C(k)$ representa al conjunto de todas las posibles combinaciones de k números naturales.

Recuérdese aquí que para Borel $P(A_k) = 0$ significa únicamente que la probabilidad de que se produzcan exactamente k éxitos tiende hacia cero cuando el número de ensayos aumenta indefinidamente. Sin embargo, en lo que sigue trataba a $P(A_0)$ como una verdadera probabilidad.

Para atacar el problema principal que se planteó Borel, definamos el evento:

A_∞ : Se obtiene éxito una infinidad de veces.

En el caso convergente, escribió Borel:

$$\begin{aligned} P(A_\infty) &= 1 - \sum_{k=0}^{\infty} P(A_k) = 1 - a(1 + \sum u_j + \sum u_{j_1} u_{j_2} + \dots) = 1 - a(1 - u_1)(1 - u_2) \dots \\ &= 1 - a \prod_{j=1}^{\infty} (1 + u_j) = 0 \end{aligned}$$

En el caso divergente, Borel se rehusaba a argumentar de la misma manera, en cuyo caso se tendría $P(A_\infty) = 1 - \sum_{k=0}^{\infty} P(A_k) = 1$.

Borel estaba rechazando aquí la propiedad de σ -aditividad.

El razonamiento de Borel era el siguiente:

“Siendo nula cada una de las probabilidades $P(A_k)$, se puede deducir que su suma también lo es y que por consiguiente $P(A_\infty)$ es igual a la unidad. El resultado es exacto, pero el razonamiento precedente carece de rigor por las razones ya indicadas. Por otra parte, es claro que no se puede buscar aquí la probabilidad de que el caso favorable se produzca una infinidad de veces en n ensayos y enseguida hacer crecer n indefinidamente; por lo tanto se razonará como sigue: eligiendo un número fijo m , se buscará la probabilidad de que el caso favorable se produzca más de m veces en n ensayos y se calculará el límite hacia el cual tiende esta probabilidad cuando n aumenta indefinidamente; omito aquí el sencillo cálculo, cuyo resultado es el siguiente: este límite es la unidad cualquiera que sea el número fijo m ; eso significa que se puede apostar con ventaja una cantidad tan grande como se quiera contra 1 franco a que el número de casos favorables será superior a un número fijo dado cualquiera m ; es precisamente la significación de este enunciado: la probabilidad $P(A_\infty)$ es igual a uno”.

Para precisar el argumento de Borel, definamos los eventos:

D_m : Se producen más de m éxitos en la sucesión infinita de ensayos.

D_m^n : Se producen más de m éxitos en los primeros n ensayos.

Tanto en el caso convergente como en el divergente, se tiene:

$$P(D_m^n) = 1 - [P(A_0^n) + \cdots + P(A_m^n)]$$

Así que:

$$P(D_m) = \lim_{n \rightsquigarrow \infty} P(D_m^n) = 1 - [P(A_0) + \cdots + P(A_m)]$$

Por lo tanto:

$$P(A_\infty) = \lim_{m \rightsquigarrow \infty} P(D_m) = \begin{cases} 0 & \text{si } \sum_{j=1}^{\infty} p_j < \infty \\ 1 & \text{si } \sum_{j=1}^{\infty} p_j = \infty \end{cases}$$

Obsérvese que en el caso divergente se tiene $P(D_m) = \lim_{n \rightsquigarrow \infty} P(D_m^n) = 1$, de manera que siendo consecuentes con el planteamiento de Borel, ésta no es una verdadera probabilidad, sino únicamente un límite, sin embargo en la expresión $P(A_\infty) = \lim_{m \rightsquigarrow \infty} P(D_m)$, Borel tomaba $P(D_m)$ como una verdadera probabilidad, lo cual, limitándonos al tipo de argumentos que da Borel, no es justificable de ninguna manera pues la probabilidad $P(A_\infty)$ no puede expresarse como un límite cuando el número de ensayos aumenta indefinidamente.

En resumen, definiendo los eventos:

A_k : Se obtienen exactamente k éxitos en la infinidad de ensayos.

A_k^n : Se producen exactamente k éxitos en los primeros n ensayos.

D_m : Se producen más de m éxitos en la sucesión infinita de ensayos.

D_m^n : Se producen más de m éxitos en los primeros n ensayos.

A_∞ : Se obtiene éxito una infinidad de veces.

Borel obtuvo los siguientes resultados:

$$P(A_k) = \lim_{n \rightsquigarrow \infty} P(A_k^n) = \begin{cases} a \sum_{\{(j_1, \dots, j_k) \in C(k)\}} u_{j_1} \cdots u_{j_k} & \text{si } \sum_{j=1}^{\infty} p_j < \infty \\ 0 & \text{si } \sum_{j=1}^{\infty} p_j = \infty \end{cases}$$

$$P(D_m) = \lim_{n \rightsquigarrow \infty} P(D_m^n) = 1 - [P(A_0) + \cdots + P(A_m)]$$

$$P(A_\infty) = \lim_{m \rightsquigarrow \infty} P(D_m) = \begin{cases} 0 & \text{si } \sum_{j=1}^{\infty} p_j < \infty \\ 1 & \text{si } \sum_{j=1}^{\infty} p_j = \infty \end{cases}$$

En realidad $P(A_k)$ ($k \in \{0, 1, \dots\}$), $P(D_m)$ ($m \in \mathbb{N}$) y $P(A_\infty)$ se obtienen asumiendo una propiedad de continuidad de la función de probabilidad. En efecto, supongamos que la función de probabilidad tiene las siguientes propiedades:

- (i) $P(\bigcup_{n=1}^{\infty} B_n) = \lim_{n \rightarrow \infty} P(B_n)$ para cualquier sucesión creciente de eventos $\{B_n\}$.
- (ii) $P(\bigcap_{n=1}^{\infty} C_n) = \lim_{n \rightarrow \infty} P(C_n)$ para cualquier sucesión decreciente de eventos $\{C_n\}$.

Entonces, se tiene:

$$A_k = \bigcap_{n=1}^{\infty} E_k^n = \bigcup_{n=1}^{\infty} F_k^n$$

en donde $E_k^n = \bigcup_{m=n}^{\infty} A_k^m$ y $F_k^n = \bigcap_{m=n}^{\infty} A_k^m$.

Ahora bien, fijando k , la sucesión $\{E_k^n\}$ es decreciente, mientras que la sucesión $\{F_k^n\}$ es creciente. Así que:

$$P(A_k) = \lim_{n \rightarrow \infty} P(E_k^n)$$

$$P(A_k) = \lim_{n \rightarrow \infty} P(F_k^n)$$

Además, $P(E_k^n) = P(\bigcup_{m=n}^{\infty} A_k^m) \geq P(A_k^n)$ y $P(F_k^n) = P(\bigcap_{m=n}^{\infty} A_k^m) \leq P(A_k^n)$ para cualquier $n \in \mathbb{N}$, así que:

$$P(A_k) = \lim_{n \rightarrow \infty} P(E_k^n) \geq \limsup_{n \rightarrow \infty} P(A_k^n)$$

$$P(A_k) = \lim_{n \rightarrow \infty} P(F_k^n) \leq \liminf_{n \rightarrow \infty} P(A_k^n)$$

Por lo tanto:

$$P(A_k) \leq \liminf_{n \rightarrow \infty} P(A_k^n) \leq \limsup_{n \rightarrow \infty} P(A_k^n) \leq P(A_k)$$

Así que la sucesión $\{P(A_k^n)\}_{n \in \mathbb{N}}$ es convergente y $P(A_k) = \lim_{n \rightarrow \infty} P(A_k^n)$.

Por otra parte:

$$D_m = \bigcup_{n=1}^{\infty} D_m^n$$

$$A_{\infty} = \bigcap_{m=1}^{\infty} D_m$$

Además, la sucesión $\{D_m\}$ es decreciente, mientras que, fijando m , la sucesión $\{D_m^n\}$ es creciente. Así que:

$$P(D_m) = \lim_{n \rightarrow \infty} P(D_m^n)$$

$$P(A_{\infty}) = \lim_{m \rightarrow \infty} P(D_m)$$

Obsérvese que las relaciones $P(A_k) = \lim_{n \rightarrow \infty} P(A_k^n)$ y $P(D_m) = \lim_{n \rightarrow \infty} P(D_m^n)$ difieren de la relación $P(A_{\infty}) = \lim_{m \rightarrow \infty} P(D_m)$ en un aspecto fundamentalmente

importante. Las dos primeras se refieren a la continuidad de la función de probabilidad cuando el número de ensayos tiende a infinito. En cambio, la última relación se refiere a la continuidad de la función de probabilidad en el marco de la sucesión infinita de ensayos. De manera que es básicamente ésta la única que rebasa el marco de la formulación clásica de la Teoría de la Probabilidad.

Los planteamientos de Borel dejan ver que para él la σ -aditividad no era una propiedad de la función de probabilidad que pueda considerarse como válida en general. Es posible que ya desde la escritura de su artículo en 1909 Borel tuviera en mente el siguiente ejemplo, el cual cita en su libro, “Applications a l’Arithmétique et a la Théorie des Fonctions”, publicado en 1926 ([18]) y en donde la σ -aditividad no se cumple: “Supongamos, por ejemplo, que existe una manera de elegir de entre la colección infinita de números enteros, uno de ellos al azar, de manera que cada uno de ellos tenga la misma probabilidad, esta probabilidad deberá entonces ser nula, pero su suma debe ser igual a 1”.

Por otra parte al asumir como válidas las relaciones $P(A_k) = \lim_{n \rightarrow \infty} P(A_k^n)$ y $P(D_m) = \lim_{n \rightarrow \infty} P(D_m^n)$ en el caso convergente y la relación $P(A_\infty) = \lim_{m \rightarrow \infty} P(D_m)$ en los dos casos, Borel estaba asumiendo implícitamente la validez de la propiedad de continuidad de la función de probabilidad mencionada antes.

Ahora bien, actualmente es bien sabido (aunque aún no lo era cuando Borel escribió su artículo) que tal propiedad de continuidad es equivalente a la σ -aditividad. De manera que los resultados de Borel asumen implícitamente que la función de probabilidad es una función σ -aditiva, es decir, una medida. De esta forma, los resultados de Borel conducen al reencuentro (paradójicamente rechazándola) con la propiedad de aditividad numerable de la función de probabilidad.

Todavía en su libro “Principios y fórmulas clásicas del Cálculo de Probabilidades”, publicado en 1925 ([14]), el análisis de Borel sobre los problemas de probabilidades numerables es prácticamente el mismo que el realizado en su artículo de 1909, no aceptando la σ -aditividad como una propiedad general de la función de probabilidad. En ese libro incluso consideró el problema de los tres jugadores (problema 1), resolviéndolo primero con el mismo método que utilizó Poincaré en su libro, es decir aplicando las reglas de la suma (principio de las probabilidades totales) y del producto (principio de las probabilidades compuestas), obteniendo así $P(A) = P(B) = \frac{5}{14}$ y $P(C) = \frac{2}{7}$, en donde A , B y C son los eventos consistentes en que P, Q y R ganen el juego respectivamente, en donde a su vez P, Q y R son los tres jugadores y se supone que se inicia el juego compitiendo P contra Q. Aclara en seguida que las probabilidades encontradas no son cocientes entre el número de casos favorables y número de casos posibles sino sumas de progresiones geométricas, las cuales obtiene de la siguiente manera:

Considerando únicamente las primeras $3n$ partidas, el jugador P únicamente puede ganar el juego en las partidas de rango $2, 4, 5, 7, \dots, 3n-2, 3n-1$, de manera que la probabilidad, $p_n(A)$, de que P gane el juego en algunas de las primeras $3n$ partidas está dada por:

$$\begin{aligned} p_n(A) &= \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^7} + \dots + \frac{1}{2^{3n-2}} + \frac{1}{2^{3n-1}} \\ &= \frac{1}{4} + \frac{3}{2^5} \left(1 + \frac{1}{2^3} + \frac{1}{2^6} + \dots + \frac{1}{2^{3n-6}} \right) = \frac{1}{4} + \frac{3}{2^5} \frac{1 - \frac{1}{2^{3n-3}}}{1 - \frac{1}{2^3}} = \frac{5}{14} - \frac{6}{7} \frac{1}{2^{3n}} \end{aligned}$$

De la misma manera, la probabilidad, $p_n(B)$, de que Q gane el juego en algunas de las primeras $3n$ partidas está dada por la misma expresión, mientras que la probabilidad, $p_n(C)$, de que R lo gane está dada por:

$$\begin{aligned} p_n(C) &= 2 \left(\frac{1}{2^3} + \frac{1}{2^6} + \frac{1}{2^9} + \dots + \frac{1}{2^{3n}} \right) \\ &= \frac{1}{4} \left(1 + \frac{1}{2^3} + \frac{1}{2^6} + \dots + \frac{1}{2^{3n-3}} \right) = \frac{1}{4} \frac{1 - \frac{1}{2^{3n}}}{1 - \frac{1}{2^3}} = \frac{2}{7} - \frac{2}{7} \frac{1}{2^{3n}} \end{aligned}$$

De manera que:

$$P(B) = P(A) = \lim_{n \rightarrow \infty} p_n(A) = \frac{5}{14}$$

$$P(C) = \lim_{n \rightarrow \infty} p_n(C) = \frac{2}{7}$$

Continuaba Borel diciendo que, de acuerdo a los cálculos anteriores, la probabilidad q_n de que el juego no se termine en las primeras $3n$ partidas está dada por:

$$q_n = 1 - 2p_n(A) - p_n(C) = \frac{12}{7} \frac{1}{2^{3n}} + \frac{2}{7} \frac{1}{2^{3n}} = \frac{1}{2^{3n-1}},$$

así que la probabilidad q_∞ de que el juego no se termine nunca, es nula.

Remarcaba entonces que “decir que la probabilidad de que el juego no se detendrá nunca es nula expresa que $\frac{1}{2^{3n-1}}$ tiende hacia cero cuando n aumenta indefinidamente”.

Este comentario de Borel deja ver que, como en los problemas de probabilidades numerables, $P(A)$, $P(B)$, $P(C)$ y q_∞ no son para él verdaderas probabilidades sino únicamente límites.

En su formulación moderna, los resultados de Borel se resumen en los siguientes dos lemas:

LEMA 7.8 (Lema de Borel Cantelli-1a. parte). *Sea A_1, A_2, \dots una sucesión de eventos tales que $\sum_{n=1}^{\infty} P(A_n) < \infty$ y sea:*

$$A = \{\omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n\}$$

Entonces $P(A) = 0$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcup_{n=m}^{\infty} A_n$. Entonces la sucesión de eventos B_m es monótona decreciente y $A = \bigcap_{m=1}^{\infty} B_m$, así que:

$$P(A) = P\left[\bigcap_{m=1}^{\infty} B_m\right] = \lim_{m \rightarrow \infty} P[B_m]$$

$$\text{Pero, } P(B_m) = P\left(\bigcup_{n=m}^{\infty} A_n\right) \leq \sum_{n=m}^{\infty} P(A_n).$$

$$\text{Por lo tanto, } P(A) \leq \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} P(A_n) = 0. \quad \blacksquare$$

LEMA 7.9 (Lema de Borel Cantelli-2a. parte). Sea A_1, A_2, \dots una sucesión de eventos independientes tales que $\sum_{n=1}^{\infty} P(A_n) = \infty$ y sea:

$$A = \{\omega \in \Omega : \omega \in A_n \text{ para una infinidad de valores de } n\}$$

entonces $P(A) = 1$.

Demostración

Para cada $m \in \mathbb{N}$, sea $B_m = \bigcap_{n=m}^{\infty} A_n^c$. Entonces la sucesión de eventos B_m es monótona creciente y $A^c = \bigcup_{m=1}^{\infty} B_m$, así que:

$$P(A^c) = P\left[\bigcup_{m=1}^{\infty} B_m\right] = \lim_{m \rightarrow \infty} P[B_m]$$

Pero, $B_m \subset \bigcap_{n=m}^{m+k} A_n^c$ para cualquier $k \in \mathbb{N}$, así que, $P(B_m) \leq \prod_{n=m}^{m+k} [1 - P(A_n)]$ para cualquier $k \in \mathbb{N}$. Por lo tanto:

$$P(B_m) \leq \lim_{k \rightarrow \infty} \prod_{n=m}^{m+k} [1 - P(A_n)] = 0$$

Se concluye entonces que $P(A) = 1 - P(A^c) = 1$. ■

7.2.1. Teorema de Borel sobre los números normales. El artículo de Borel causó un gran impacto en su época sobre todo por una aplicación de sus resultados para deducir una propiedad importante de los números reales, la cual se expone en esta parte.

Si q es un número natural mayor que 1, llamemos una fracción decimal de base q a una expresión de la forma $\sum_{j=1}^{\infty} \frac{b_j}{q^j}$, en donde cada b_j es un entero no negativo menor que q . Supongamos además que:

- (i) Cada número b_j se elige de tal manera que la probabilidad de que tome cada uno de los valores $0, \dots, q-1$ es igual a $\frac{1}{q}$.
- (ii) Las elecciones de los números b_j son independientes unas de otras.

Dado un número $b \in \{0, \dots, q-1\}$ y una fracción decimal $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$, se define la frecuencia de b en x hasta el rango n , $f_n(b)$, como el cociente que resulta de dividir el número de veces que aparece b en los primeros n términos de x entre n . Cuando $\lim_{n \rightarrow \infty} f_n(b)$ existe, se dirá que la frecuencia total de b en x existe y que su valor es igual a ese límite.

Se dice que la fracción decimal $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$ es simplemente normal con respecto a la base q si dado cualquier número $b \in \{0, \dots, q-1\}$, la frecuencia total de b en x existe y su valor es igual a $\frac{1}{q}$.

Borel demostró entonces que la probabilidad de que un número $x = \sum_{j=1}^{\infty} \frac{b_j}{q^j}$ sea simplemente normal con respecto a la base q es igual a 1. Para la base $q = 2$, su demostración es como sigue:

La formación de un número $x = \sum_{j=1}^{\infty} \frac{b_j}{2^j}$ se puede realizar efectuando una sucesión de ensayos independientes en cada uno de los cuales la probabilidad de obtener éxito es igual a la de obtener fracaso, es decir $\frac{1}{2}$, de tal manera que cuando en el ensayo j se obtiene éxito se define $b_j = 0$, mientras que cuando se obtiene fracaso se define $b_j = 1$. Para cada $n \in \mathbb{N}$, definamos:

S_n : número de éxitos hasta el ensayo n .

R_n : número de fracasos hasta el ensayo n .

$$B_n = \left[\left| S_n - \frac{n}{2} \right| > \frac{\sqrt{n \ln n}}{\sqrt{2}} \right]$$

Entonces, aplicando el teorema de de Moivre-Laplace, se puede demostrar que $\sum_{n=1}^{\infty} P(B_n) < \infty$.

Si $B = \left\{ \left| S_n - \frac{n}{2} \right| > \frac{\sqrt{n \ln n}}{\sqrt{2}} \text{ para una infinidad de valores de } n \right\}$, entonces se tiene, por el resultado de Borel, $P(B) = 0$.³

³El argumento de Borel no es totalmente correcto pues para probar $P(B) = 1$, asume que los eventos B_n son independientes, lo cual no es cierto en este caso. Sin embargo, se puede aplicar el lema 7.8, el cual no requiere de la hipótesis de independencia.

Por lo tanto, con probabilidad 1, $|S_n - \frac{n}{2}| \leq \frac{\sqrt{n \ln n}}{\sqrt{2}}$ a partir de un cierto valor de n . Así que, a partir de ese valor, se tiene $\frac{n}{2} - \frac{\sqrt{n \ln n}}{\sqrt{2}} \leq S_n \leq \frac{n}{2} + \frac{\sqrt{n \ln n}}{\sqrt{2}}$ y $\frac{n}{2} - \frac{\sqrt{n \ln n}}{\sqrt{2}} \leq R_n \leq \frac{n}{2} + \frac{\sqrt{n \ln n}}{\sqrt{2}}$, lo cual implica:

$$\frac{\frac{n}{2} - \frac{\sqrt{n \ln n}}{\sqrt{2}}}{\frac{n}{2} + \frac{\sqrt{n \ln n}}{\sqrt{2}}} \leq \frac{S_n}{R_n} \leq \frac{\frac{n}{2} + \frac{\sqrt{n \ln n}}{\sqrt{2}}}{\frac{n}{2} - \frac{\sqrt{n \ln n}}{\sqrt{2}}}$$

Es decir:

$$\frac{1 - \frac{\sqrt{2 \ln n}}{\sqrt{n}}}{1 + \frac{\sqrt{2 \ln n}}{\sqrt{n}}} \leq \frac{S_n}{R_n} \leq \frac{1 + \frac{\sqrt{2 \ln n}}{\sqrt{n}}}{1 - \frac{\sqrt{2 \ln n}}{\sqrt{n}}}$$

Así que, $P \left[\lim_{n \rightarrow \infty} \frac{S_n}{R_n} = 1 \right] = 1$, lo cual implica el resultado.

7.3. Surgimiento de la Teoría de la Medida

Con el objeto de ubicar el paralelismo que se da entre el desarrollo de la Teoría de la Medida y la Teoría de la Probabilidad a principios del siglo XX, se expone a continuación la manera en que surgió la Teoría de la Medida.

7.3.1. La integral de Cauchy. Aunque los conceptos de contenido o de medida de un conjunto pueden pensarse como una extensión de los conceptos de longitud, área, volumen, etc., en realidad, históricamente, surgieron de la Teoría de Integración.

La definición analítica de la integral de una función fue formulada por vez primera por Augustin-Louis Cauchy en el año 1823 ([30]). En ese trabajo, Cauchy definió el concepto de continuidad básicamente como se conoce actualmente:

Una función definida en un intervalo es continua si para cada x en el intervalo el valor numérico de la diferencia $f(x + \alpha) - f(x)$ decrece indefinidamente con α .

Además, formuló la definición analítica de la integral de una función continua, demostrando su existencia:

Sea f una función continua en el intervalo $[a, b]$, entonces las sumas:

$$S = \sum_{k=1}^n f(x_{k-1}) (x_k - x_{k-1}),$$

correspondientes a particiones $P = \{a = x_0 < \dots < x_n = b\}$ tienden a un límite cuando los elementos $x_k - x_{k-1}$ se hacen infinitamente pequeños; a ese límite se le llama la integral definida de f y se le denota por $\int_a^b f(x) dx$. Se obtiene el mismo límite

si se consideran sumas de la forma $S = \sum_{k=1}^n f[x_{k-1} + \theta_k(x_k - x_{k-1})](x_k - x_{k-1})$, en donde $0 \leq \theta_k \leq 1$.

Demostró además que si f es una función continua y $F(x) = \int_a^x f(y)dy$, entonces $F'(x_0) = f(x_0)$.

La integral así definida es conocida actualmente como la integral de Riemann y no como la integral de Cauchy. La razón de esto parece justa pues es el trabajo de Georg Friedrich Bernhard Riemann, publicado en el año 1867, el que dió la pauta para desarrollar una Teoría de Integración, la cual a su vez llevaría más tarde a una Teoría del Contenido y finalmente a la moderna Teoría de la Medida.

En trabajos posteriores, Cauchy consideró funciones discontinuas haciendo la aclaración siguiente:

“es necesario observar que las funciones discontinuas introducidas en el Cálculo dejan de ser continuas únicamente para algunos valores de las variables”

Para este tipo de funciones discontinuas extendió el concepto de integral de la siguiente manera:

Si una función es continua en un intervalo $[a, b]$, excepto en un punto c , en una vecindad del cual f puede ser acotada o no, se puede definir la integral de f como el límite:

$$\lim_{h \rightarrow 0} \left[\int_a^{c-h} f(x)dx + \int_{c-h}^b f(x)dx \right]$$

cuando éste existe.

Entre 1822 y 1825, Johann Peter Gustav Lejeune Dirichlet extendió el trabajo de Cauchy considerando funciones que admiten un número finito de discontinuidades. Conjeturó más adelante (1829) que los métodos de Cauchy, incluyendo la existencia de la integral, se pueden extender a todas las funciones que tengan la siguiente propiedad:

Suponiendo que f está definida en un intervalo $[a, b]$, dadas dos cantidades arbitrarias u y v en ese intervalo, es posible encontrar otras dos cantidades r y s entre u y v tales que la función f es continua en el intervalo $[r, s]$.

Es decir, utilizando la terminología moderna, el conjunto de puntos en donde la función es discontinua debe ser denso en ninguna parte.

En 1864, Rudolf Otto Sigismund Lipschitz ([77]) “demostró” la conjetura de Dirichlet. Para esto daba por hecho que la condición de Dirichlet implica que el conjunto de

puntos de acumulación del conjunto de discontinuidades de f debe de ser finito. En otras palabras, asumía que la condición de Dirichlet implica que las discontinuidades de f se acumulan alrededor de un número finito de puntos.

Si f es discontinua únicamente en un número finito de puntos y c es un punto en donde f es discontinua, Lipschitz definía:

$$\int_a^b f(x)dx = \lim_{\varepsilon \rightarrow 0} \left[\int_a^{c-\varepsilon} f(x)dx + \int_{c+\varepsilon}^b f(x)dx \right]$$

Si c es un punto de acumulación del conjunto de discontinuidades de f , definía:

$$\int_a^b f(x)dx = \lim_{\varepsilon \rightarrow 0} \left[\int_a^{c-\varepsilon} f(x)dx + \int_{c+\varepsilon}^b f(x)dx \right]$$

La identificación que hacía Lipschitz de los conjuntos densos en ninguna parte con los conjuntos que tienen un número finito de puntos de acumulación es errónea. En efecto, si un conjunto no es denso en ninguna parte entonces existe un intervalo en el cual es denso y, por lo tanto, todos los puntos de ese intervalo son puntos de acumulación del conjunto, es decir, si un conjunto no es denso en ninguna parte entonces el conjunto de sus puntos de acumulación es infinito. Sin embargo, el inverso de este resultado no es válido; en efecto, para cada $n \in \mathbb{N}$, consideremos una sucesión decreciente de puntos aislados del intervalo $(\frac{1}{n}, \frac{1}{n-1})$ que converja a $x_n = \frac{1}{n}$. El conjunto así formado es denso en ninguna parte pero tiene una infinidad de puntos de acumulación.

Durante un tiempo prevaleció esta idea errónea, según la cual los conjuntos despreciables para la Teoría de la Integración son los conjuntos densos en ninguna parte, los cuales a su vez, también erróneamente, eran identificados con aquellos cuyo conjunto de puntos de acumulación es finito.

Los trabajos sobre el concepto de integral previos al de Riemann muestran que lo que se buscaba era extender la definición de la integral a funciones tan discontinuas como fuera posible. La definición analítica de la integral no era la misma para las funciones continuas que para las discontinuas.

7.3.2. La integral de Riemann. Georg Friedrich Bernhard Riemann, en un artículo, elaborado en 1854 y publicado por Julius Wilhelm Richard Dedekind en 1867 ([91]), cambió el enfoque para atacar el problema de la integración de funciones. Como lo mencionamos antes, antes de él se trataba de extender la definición de la integral a funciones que tuvieran tantas discontinuidades como fuera posible. Para Riemann, la integral de cualquier función acotada definida en un intervalo cerrado debía definirse esencialmente como lo hizo Cauchy para las funciones continuas:

Consideremos una partición x_0, \dots, x_n del intervalo $[a, b]$ y definamos $\delta_k = x_k - x_{k-1}$; si, independientemente de como se elijan las cantidades $\varepsilon_k \in [0, 1]$, las sumas $\sum_{k=1}^n \delta_k f(x_{k-1} + \varepsilon_k \delta_k)$ tienden a un límite cuando todas las cantidades $\delta_k = x_k - x_{k-1}$ tienden a cero, a ese límite se le llama el valor de la integral definida $\int_a^b f(x)dx$.

Aclaraba Riemann que cuando el límite de tales sumas no existe entonces la notación $\int_a^b f(x)dx$ carece de significado.

Una vez establecida la definición, Riemann se planteó el problema de caracterizar a aquellas funciones para las cuales el límite que define la integral existe:

“Busquemos ahora la extensión y el límite de la definición precedente y hagámonos esta pregunta: ¿en qué casos una función es susceptible de integración?, ¿en qué casos no lo es?”

Estableció dos criterios, ambos basados en el concepto de oscilación de una función en un intervalo.

DEFINICIÓN 7.10 (Oscilación de una función en un intervalo). Sea $f : [a, b] \mapsto \mathbb{R}$ una función acotada. La diferencia:

$$\sup \{f(x) : x \in [x_{k-1}, x_k]\} - \inf \{f(x) : x \in [x_{k-1}, x_k]\}$$

es llamada la oscilación de f en el intervalo $[a, b]$.

Criterio R_1

Sea D_k la oscilación de f en el intervalo $[x_{k-1}, x_k]$, entonces:

$$f \text{ es integrable si y sólo si } \lim_{\delta_k \rightsquigarrow 0} \sum_k D_k \delta_k = 0$$

Este criterio, conocido simplemente como el criterio de Riemann, se formula y demuestra actualmente en casi cualquier libro de Análisis Matemático.

Criterio R_2

Dada $\sigma > 0$ y una partición P , sea $\lambda(P, \sigma)$ la suma de las longitudes de los subintervalos de la partición en los cuales la oscilación de la función es mayor que σ , entonces:

$$f \text{ es integrable si y sólo si } \lim_{\|P\| \rightsquigarrow 0} \lambda(P, \sigma) = 0 \quad \forall \sigma > 0$$

en donde $\|P\|$ es la norma de P .

Este criterio se sigue del criterio R_1 y las siguientes desigualdades:

$$\sigma\lambda(P, \sigma) \leq \sum_k D_k \delta_k \leq D\lambda(P, \sigma) + (b - a)\sigma$$

en donde D_k es la oscilación de f en el intervalo $[x_{k-1}, x_k]$ y D la oscilación de f en el intervalo $[a, b]$.

El criterio R_2 permitió a Riemann dar un ejemplo de una función integrable con un conjunto denso de discontinuidades:

Sea $M = \{\frac{1}{2}, \frac{3}{2}, \dots\}$ y, para $x \in [0, \infty)$, definamos:

$$f(x) = \begin{cases} 0 & \text{si } x \in M \\ x - m(x) & \text{si } x \notin M \end{cases}$$

en donde $m(x)$ es el número entero más cercano a x .

Riemann definió entonces la función $f : [0, 1] \mapsto \mathbb{R}$ de la siguiente manera:

$$f(x) = \sum_{k=1}^{\infty} \frac{(kx)}{k^2}$$

Se puede demostrar que esta función es discontinua en todos los puntos x de la forma $x = \frac{m}{2n}$, en donde m y n son dos números naturales tales que m y $2n$ son primos entre sí. Además f satisface el criterio R_2 de Riemann y, por lo tanto, es integrable.

7.3.3. De la Teoría de Integración a la Teoría del Contenido. Aún después de conocerse el trabajo de Riemann se siguió pensando que si el conjunto de puntos de discontinuidad de una función es denso en ninguna parte, entonces la función es integrable, es decir se pensaba que la condición de Dirichlet es más restrictiva que la de Riemann.

Hermann Hankel, discípulo de Riemann, introdujo en 1870 ([53]) el concepto de oscilación de una función en un punto y reformuló el criterio de Riemann en los siguientes términos:

Sea $f : [a, b] \mapsto \mathbb{R}$ una función acotada y $x \in (a, b)$. Sea (I_n) una sucesión de intervalos cerrados encajados que contengan a x como punto interior y tales que $\cap I_n = \{x\}$; denotemos por O_n a la oscilación de f en el intervalo I_n ; entonces el $\lim O_n$ existe y es independiente de la sucesión particular de intervalos encajados con las propiedades dadas antes. A ese límite se le llama la **oscilación de la función f en el punto x** .

Demostó entonces, erróneamente, que una función es integrable si y sólo si para cualquier $\varepsilon > 0$ el conjunto de puntos en donde la oscilación de la función es mayor que ε es denso en ninguna parte.

Durante varios años prevaleció la búsqueda de la caracterización de las funciones integrables con base en la pequeñez topológica del conjunto de sus discontinuidades y, en esta búsqueda, se puede observar la confusión que existía respecto a los diferentes conceptos de pequeñez que podían definirse. Alrededor del año 1873 tal confusión radicaba básicamente en la idea de que un conjunto es denso en ninguna parte si y sólo si es de primera especie.

Si $A \subset \mathbb{R}$, se denota por $A^{(1)}$ al conjunto de puntos de acumulación de A , por $A^{(2)}$ al conjunto de puntos de acumulación de $A^{(1)}$, etc... Al conjunto $A^{(n)}$ se le llama el n -ésimo conjunto derivado de A . Se dice que un conjunto $A \subset \mathbb{R}$ es de primera especie si $A^{(n)}$ es finito para alguna n .

En 1873 era ya bien conocido que un conjunto acotado de primera especie es denso en ninguna parte: si un conjunto es denso en algún intervalo, entonces el conjunto de sus puntos de acumulación también lo es; de manera que ese conjunto no puede ser de primera especie

Sin embargo, se pensaba que los conjuntos de primera especie agotaban las posibilidades de los conjuntos densos en ninguna parte. La confusión terminó cuando se inventaron métodos para construir conjuntos densos en ninguna parte.

Paul David Gustav du Bois-Reymond dio en 1880 un ejemplo de un conjunto denso en ninguna parte que no es de primera especie:

Sea I_n una sucesión de intervalos ajenos cuyos puntos extremos convergen al punto P .

En el interior de I_n definamos un conjunto Q_n de orden n y sea $Q = \bigcup_n Q_n$.

Q es un conjunto denso en ninguna parte pues cada conjunto Q_n lo es y éstos se encuentran en intervalos ajenos.

Por otra parte, $P \in Q^{(n)}$ para toda n , por lo tanto, Q no es de primera especie.

Otro método de construcción de conjuntos densos en ninguna parte fue desarrollado de manera independiente por Henry John Stephen Smith en 1875 ([93]), Vito Volterra en 1881 ([97], [98]) y Georg Ferdinand Ludwig Philipp Cantor durante el periodo 1879-1884 ([23], [24], [25], [26], [27]). Este método es el que se utiliza actualmente para definir el conjunto de Cantor, el cual es un ejemplo de un conjunto denso en ninguna parte que no es de primera especie.

Definamos:

$$F_0 = [0, 1]$$

$$F_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$$

$$F_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$$

⋮

En general, si ya tenemos definido el conjunto F_n , éste consta de una unión de 2^n intervalos cerrados ajenos. El conjunto F_{n+1} se construye entonces partiendo cada uno de esos intervalos en 3 intervalos de la misma longitud y eliminando el intervalo central abierto.

$F = \cap F_n$ es llamado el conjunto de Cantor y tiene las siguientes propiedades:

- Es un conjunto denso en ninguna parte.
- $F = F^{(n)}$ para toda n , por lo tanto, no es de primera especie.

Durante este periodo emergió una nueva clase de conjuntos, los de contenido cero:

DEFINICIÓN 7.11 (Conjuntos de contenido cero). *Se dice que un conjunto tiene contenido cero si, para cualquier $\varepsilon > 0$, existe una familia finita de intervalos abiertos que cubren al conjunto y tales que la suma de sus longitudes es menor que ε .*

Se pudo demostrar además que esta nueva clase de conjuntos se ubica entre las otras dos que hemos mencionado, es decir, todo conjunto acotado de primera especie tiene contenido cero y a su vez todo conjunto de contenido cero es denso en ninguna parte.

Todo lo anterior permitió exhibir funciones no integrables cuyo conjunto de discontinuidades sea denso en ninguna parte: La función indicadora de un conjunto denso en ninguna parte que no tiene contenido cero tiene un conjunto de discontinuidades denso en ninguna parte, pero no es integrable.

Por otro lado, Riemann había mostrado la existencia de funciones cuyas discontinuidades forman un conjunto denso pero que son integrables. Se podía concluir, finalmente, que **no es el tamaño topológico del conjunto de discontinuidades lo que determina que una función sea o no sea integrable.**

Es en ese momento cuando se pudo ya establecer con toda claridad la condición para que una función sea integrable. Axel Harnack demostró en 1881 que:

Una función es integrable si y sólo si, para cualquier $\sigma > 0$, el conjunto de puntos en donde la oscilación de la función es mayor que σ tiene contenido cero.([54])

El concepto de contenido cero se convertiría desde ese momento en uno clave para la Teoría de la Integración.

En ese momento se tuvieron entonces las bases para desarrollar una Teoría del Contenido, lo cual fue llevado a cabo por Otto Stolz ([94]), Axel Harnack ([55], [56]), Giuseppe Peano ([87])y, sobre todo, Marie Ennemond Camille Jordan ([60]). Todo esto durante el periodo que va de 1883 a 1892:

Sea A un conjunto acotado de números reales y $[a, b]$ un intervalo que lo contenga. Para cada partición P del intervalo $[a, b]$ sea $\bar{S}(P, A)$ la suma de los subintervalos de P que contienen puntos de A y $\underline{S}(P, A)$ la suma de los subintervalos de P contenidos en A . Se define entonces el **contenido exterior** de A , $c_e(A)$, y el **contenido interior** de A , $c_i(A)$ mediante las relaciones:

$$c_e(A) = \inf \{ \bar{S}(P, A) : P \text{ es partición del intervalo } [a, b] \}$$

$$c_i(A) = \inf \{ \underline{S}(P, A) : P \text{ es partición del intervalo } [a, b] \}$$

Se dice entonces que A es **Jordan-medible** si $c_e(A) = c_i(A)$ y, en este caso, a esta cantidad común se le llama el **contenido** de A y se le denota por $c(A)$.

Evidentemente todo conjunto de contenido cero es Jordan-medible. También todo intervalo acotado es Jordan-medible y su contenido es igual a su longitud. Finalmente, se tiene la siguiente propiedad:

Consideremos un intervalo $[a, b]$, entonces la familia de subconjuntos de $[a, b]$ que son Jordan-medibles forman un álgebra de subconjuntos de $[a, b]$. Además, la función que asocia a cada conjunto Jordan-medible su contenido es finitamente aditiva.

7.3.4. Teoría de la Medida de Borel. En 1894-1895, Félix Édouard Justin Émile Borel ([12], [13]) dio las bases para un nuevo avance al introducir el concepto de medida cero:

Se dice que un conjunto tiene **medida cero** si para cualquier $\varepsilon > 0$ existe una colección numerable de intervalos abiertos $\{I_n\}$ tales que la suma de sus longitudes es menor que ε .

Curiosamente, el concepto de medida cero no lo introdujo Borel con relación a la Teoría de Integración. Al introducir ese concepto, Borel estaba atacando un problema de continuación analítica de una función de variable compleja:

Considérese la función de variable compleja:

$$f(z) = \sum_{n=1}^{\infty} \frac{A_n}{z-a_n}$$

en donde A_1, A_2, \dots son números complejos tales que la serie $\sum_{n=1}^{\infty} |A_n|$ converge y a_1, a_2, \dots son puntos en el plano complejo que están sobre una curva cerrada C formando un conjunto denso en esa curva.

Se puede ver inmediatamente que si $z \notin C$ entonces la serie $\sum_{n=1}^{\infty} \frac{A_n}{z-a_n}$ converge pues la distancia de z a C es positiva. Consideremos dos puntos P y Q , el primero al interior de la región que forma C y el segundo al exterior de la misma; el problema que se plantea Borel consiste entonces en encontrar un arco circular que una P con Q sobre el cual la serie $\sum_{n=1}^{\infty} \frac{A_n}{z-a_n}$ converja absoluta y uniformemente. Esto llevó a Borel a la necesidad de demostrar que existen puntos z sobre C para los cuales la serie en consideración converge.

Para simplificar el razonamiento, consideremos el mismo problema pero con funciones de variable real.

Sea $\{a_1, a_2, \dots\}$ un conjunto numerable y denso en el intervalo $[a, b]$ y $(A_n)_{n \geq 1}$ una sucesión de números reales. Para cada $x \in [a, b] - \{a_1, a_2, \dots\}$ consideremos la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$. Aparentemente tal serie no converge para ninguno de esos puntos x pues el conjunto $\{a_1, a_2, \dots\}$ es denso en $[a, b]$ y entonces dado cualquier punto $x \in [a, b]$ hay puntos a_n tan cerca de x como se quiera. Sin embargo, siguiendo a Borel, se puede mostrar que, asumiendo que la serie $\sum_{n=1}^{\infty} \sqrt{|A_n|}$ converge, existe una infinidad no numerable de puntos $x \in [a, b]$ para los cuales la serie converge. En efecto, para cada $n \in \mathbb{N}$, sea $u_n = \sqrt{|A_n|}$. Sea ahora l la longitud del intervalo $[a, b]$ y $N \in \mathbb{N}$ tal que $\sum_{n=N+1}^{\infty} u_n < \frac{l}{2}$. Para cada $n > N$ sea I_n un intervalo abierto con centro en a_n y radio u_n . Se tiene entonces $\sum_{n=N+1}^{\infty} l(I_n) < l$, en donde $l(I_n)$ es la longitud del intervalo I_n . Como los puntos a_1, \dots, a_N forman un conjunto finito, se pueden cubrir con intervalos abiertos I_1, I_2, \dots, I_N , respectivamente, de tal manera que $\sum_{n=1}^{\infty} l(I_n) < l$. Si x no pertenece a ninguno de los intervalos I_1, I_2, \dots entonces $|x - a_i| > 0$ para $i \in \{1, \dots, N\}$ y $|x - a_i| \geq u_i$ para $i \in \{N+1, N+2, \dots\}$. Por lo tanto:

$$\begin{aligned} \sum_{n=1}^{\infty} \left| \frac{A_n}{x-a_n} \right| &= \sum_{n=1}^N \left| \frac{A_n}{x-a_n} \right| + \sum_{n=N+1}^{\infty} \left| \frac{A_n}{x-a_n} \right| \\ &\leq \sum_{n=1}^N \left| \frac{A_n}{x-a_n} \right| + \sum_{n=N+1}^{\infty} \sqrt{|A_n|} < \infty \end{aligned}$$

Lo único que resta probar es que existe una infinidad de puntos $x \in [a, b]$ que no pertenecen a ninguno de los intervalos I_1, I_2, \dots . Para esto, Borel demostró el resultado, ahora clásico, que asegura que todo intervalo cerrado y acotado es compacto. De manera más específica, Borel demostró, básicamente como se hace actualmente,

que si un intervalo cerrado y acotado es cubierto por una infinidad numerable de intervalos abiertos, entonces existe una colección finita de esos intervalos que también lo cubren. En base a este resultado, si los intervalos I_1, I_2, \dots cubrieran al intervalo $[a, b]$, necesariamente se tendría $\sum_{n=1}^{\infty} l(I_n) \geq l$, lo cual es una contradicción. Más aún, si únicamente hubiera una colección numerable de puntos $x \in [a, b]$ que no pertenecen a ninguno de los intervalos I_1, I_2, \dots , estos puntos podrían ser cubiertos por una nueva colección numerable de intervalos abiertos de tal manera que la suma de sus longitudes, sumadas con las longitudes de los intervalos I_1, I_2, \dots , siga siendo menor que l , lo cual no es posible.

Todavía siguiendo a Borel, se puede decir aún más, pues cambiando l por una $\varepsilon > 0$ arbitraria en el razonamiento anterior se muestra que el conjunto de puntos $x \in [a, b]$ para los cuales la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$ no converge absolutamente pueden ser cubiertos por una colección numerable de intervalos abiertos de tal manera que la suma de sus longitudes sea menor que ε . Es decir, utilizando el concepto que introduce Borel, el conjunto de puntos $x \in [a, b]$ para los cuales la serie $\sum_{n=1}^{\infty} \frac{A_n}{x-a_n}$ no converge absolutamente tiene medida cero.

Más adelante, en un libro publicado en 1898 ([14]), Borel retomó el concepto de conjunto de medida cero para desarrollar una Teoría de la Medida. Para esto, influenciado en parte por el trabajo de Jules Joseph Drach, siguió el método axiomático. Para Borel la idea fundamental consistía en definir los elementos nuevos que se introducen con ayuda de sus propiedades esenciales, es decir, aquellas que son estrictamente indispensables para los razonamientos que siguen. En el caso de la medida, las propiedades esenciales que planteó Borel son las siguientes:

- (i) La medida de la unión de una colección numerable de conjuntos ajenos es igual a la suma de sus medidas.
- (ii) La medida de la diferencia de dos conjuntos de medida finita A y B , con $A \subset B$, es igual a la diferencia de sus medidas, $m(B) - m(A)$.
- (iii) La medida de un conjunto nunca es negativa.

Llamó entonces conjuntos medibles a todos aquellos conjuntos a los cuales se les pueda asignar una medida en base a las propiedades mencionadas y tomando como punto de partida que la medida de un intervalo es su longitud.

Borel no vio relación entre su concepto de medida y el de integral. Más aún, aclaraba que el problema que él estaba investigando era totalmente diferente del resuelto por Jordan. Además, consideraba la definición que hacía Jordan de los conjuntos medibles (con contenido) como más general que la que él daba pues, por ejemplo, con base en la definición de Jordan, cualquier subconjunto del conjunto de Cantor es medible, de manera que, teniendo el conjunto de Cantor la misma cardinalidad que los números

reales, la familia de conjuntos Jordan medibles tiene una cardinalidad mayor que la de los reales. Por otra parte, se puede mostrar que la familia de conjuntos medibles que define Borel tiene únicamente la cardinalidad de los números reales.

7.3.5. Teoría de la Medida de Lebesgue. El paso siguiente en el desarrollo de la teoría de la medida, así como el último paso hacia la caracterización de las funciones Riemann-integrables lo dio Henri Léon Lebesgue en 1902.

Para la caracterización de las funciones Riemann integrables, Lebesgue primero demostró una forma ligeramente distinta del resultado de Harnack:

Si, dada $\sigma > 0$, $B(\sigma)$ denota al conjunto de puntos en donde la oscilación de la función f es mayor o igual que σ , entonces f es integrable si y sólo si para cualquier $\sigma > 0$, $B(\sigma)$ tiene contenido cero.

Mostró además que, para cualquier $\sigma > 0$, $B(\sigma)$ es un conjunto cerrado, de manera que, siendo acotado, es compacto.

Lebesgue observó entonces que si D es el conjunto de puntos en donde la función es discontinua, se tiene $D = \bigcup_{n=1}^{\infty} B(\frac{1}{n})$. Entonces, si f es Riemann integrable, $B(\frac{1}{n})$ tiene contenido cero para cualquier $n \in \mathbb{N}$, así que D tiene medida cero. Por otra parte, si D tiene medida cero, entonces $B(\frac{1}{n})$ tiene medida cero para cualquier $n \in \mathbb{N}$, de manera que, siendo estos conjuntos compactos, también tienen contenido cero; finalmente, dada $\sigma > 0$ arbitraria y $n > \frac{1}{\sigma}$ se tiene $B(\sigma) \subset B(\frac{1}{n})$, así que $B(\sigma)$ tiene contenido cero. Se tiene así la siguiente caracterización de las funciones Riemann integrables:

Una función acotada $f : [a, b] \mapsto \mathbb{R}$ es Riemann integrable si y sólo si el conjunto de puntos en donde la función es discontinua tiene medida cero.

Lebesgue desarrolló su Teoría de la Medida en su tesis doctoral titulada “Integrale, longueur, aire” ([72]). Ahí, siguiendo a Borel, comenzó planteándose lo que él llamó el problema de la medida, el cual consiste en definir una medida no negativa m sobre todos los conjuntos acotados de números reales de tal manera que se tengan las siguientes propiedades:

- (i) $m([0, 1]) = 1$.
- (ii) Si E es un conjunto acotado y $a \in \mathbb{R}$, entonces $m(E + a) = m(E)$.
- (iii) Si E_1, E_2, \dots es una sucesión de conjuntos contenidos en un conjunto acotado y ajenos por parejas, entonces $m(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} m(E_k)$.

Se puede ver fácilmente que las condiciones sobre la medida implican que la medida de un intervalo acotado debe de ser igual a su longitud.

Consideró entonces el problema de la medida limitándose a subconjuntos de un intervalo fijo I .

Si E es un subconjunto de I al cual se le asigna la medida $m(E)$ e I_1, I_2, \dots es una colección finita o infinita numerable de intervalos ajenos tales que $E \subset \bigcup_n I_n$, entonces se debe de tener $m(E) \leq \sum_n l(I_n)$, de manera que la cantidad:

$$\inf \left\{ \sum_n l(I_n) : I_1, I_2, \dots \text{ son intervalos ajenos y } E \subset \bigcup_n I_n \right\}$$

es una cota superior para la medida de E . Definió entonces la **medida exterior** de E , $m_e(E)$, como esa cantidad, es decir:

$$m_e(E) = \inf \left\{ \sum_n l(I_n) : I_1, I_2, \dots \text{ son intervalos ajenos y } E \subset \bigcup_n I_n \right\}$$

En seguida definió la **medida interior** de E , $m_i(E)$, mediante la relación:

$$m_i(E) = l(I) - m_e(E^c)$$

Ahora bien, por la propiedad *iii* se debe de tener $l(I) = m(E) + m(E^c)$, es decir, $m(E) = l(I) - m(E^c)$. Pero se tiene $m(E^c) \leq m_e(E^c)$, así que $m(E) \geq l(I) - m_e(E^c)$. De esta forma se obtiene que la cantidad $l(I) - m_e(E^c)$ es una cota inferior de $m(E)$.

Todo lo anterior lo hacía Lebesgue asumiendo que es posible asignarle una medida al conjunto E . Sin embargo las definiciones de medida exterior e interior son independientes de esta consideración y pueden darse para cualquier conjunto. Mostró entonces que se tienen las siguientes relaciones para cualquier subconjunto E de I :

$$c_i(E) \leq m_i(E) \leq m(E) \leq m_e(E) \leq c_e(E)$$

Además, como se mostró arriba, de ser posible asignar una medida $m(E)$ al conjunto E , se debe de tener $m_i(E) \leq m(E) \leq m_e(E)$. Por lo tanto, la medida asignada a E será única cuando sus medidas interior y exterior coincidan.

Estas consideraciones condujeron a Lebesgue a su definición de medibilidad:

DEFINICIÓN 7.12 (Conjuntos medibles). *Se dice que un conjunto acotado E es medible si $m_i(E) = m_e(E)$.*

Aclaraba Lebesgue que es únicamente para estos conjuntos que se estudiará el problema de la medida, declarando no saber siquiera si existen conjuntos que no sean medibles. Pero si existen tales conjuntos, decía que el desarrollo posterior que él hace

no es suficiente para afirmar ni que el problema de la medida es posible ni que es imposible para tales conjuntos.

Este comentario de Lebesgue es importante pues lo que él ha hecho es encontrar cotas más finas que las que da Jordan para la medida de un conjunto, lo cual automáticamente amplía la familia de conjuntos a los cuales se les puede asignar una medida de manera única. En efecto, la condición $c_i(E) = c_e(E)$ permite asignar a E una única medida y esa condición implica $m_i(E) = m_e(E)$. Pero se puede cumplir la condición $m_i(E) = m_e(E)$, lo cual permite asignar una única medida a E , sin que se tenga $c_i(E) = c_e(E)$. Sin embargo, no se puede asegurar que no sea posible asignarle una medida a conjuntos para los cuales $m_i(E) < m_e(E)$. En caso de que esto fuera posible, tal vez no sería de manera única (de hecho se sabe actualmente que es posible ampliar la familia de conjuntos medibles conservando las propiedades *i* y *iii* que pide Lebesgue a la medida, pero tal extensión no es única), o tal vez se puedan encontrar cotas aún más finas que las que da Lebesgue para la medida de un conjunto y se pueda definir una medida con propiedades adicionales a las que propone Lebesgue.

Mostró Lebesgue que se tienen las siguientes propiedades:

- (i) Si E es medible entonces E^c es medible.
- (ii) Si A y B son medibles, entonces $A - B$ es medible.
- (iii) Si E_1, \dots, E_n son conjuntos medibles, entonces $\bigcup_{j=1}^n E_j$ es medible.
- (iv) Si E_1, E_2, \dots es una colección finita o infinita numerable de subconjuntos de I , entonces $m_e(\bigcup_n E_n) \leq \sum_n m_e(E_n)$.

Demostó además que la familia de conjuntos medibles contenidos en un intervalo I forma una σ -álgebra de subconjuntos de ese intervalo y que la medida m definida sobre esa σ -álgebra es σ -aditiva.

Finalmente observó Lebesgue que, debido a la relación $c_i(E) \leq m_i(E) \leq m_e(E) \leq c_e(E)$, cualquier conjunto Jordan medible es también Lebesgue medible y, dado que los intervalos son medibles y la familia de conjuntos medibles contenidos en un intervalo forma una σ -álgebra de subconjuntos de ese intervalo, todo conjunto medible de acuerdo a la definición de Borel es también Lebesgue medible. De esta forma la Teoría de la Medida de Lebesgue resulta más general tanto que la de Jordan como de la de Borel y las engloba a ambas.

El trabajo de Lebesgue sobre la Teoría de la Medida se puede consultar en su libro [73], el cual es una verdadera joya.

7.4. Identificación de funciones de probabilidad con medidas

Inmediatamente después del surgimiento de la Teoría de la Medida de Lebesgue se dio una relación con la Teoría de la Probabilidad. Esto se hizo en los problemas de probabilidades geométricas, en cuyo caso la probabilidad era considerada como una medida.

En 1904, Borel planteó que la integral clásica (de Riemann) es insuficiente para tratar algunos problemas de probabilidad ([15]):

Si se sabe que un número x está comprendido entre 0 y 1, ¿cuál es la probabilidad de que x sea un número racional?

Utilizando la integral de Riemann, el problema no tiene solución.

Utilizando la integral de Lebesgue, la respuesta es 0.

En 1911, Sergi Natanovich Bernstein utilizó las diferentes formas de la aditividad numerable para un problema de probabilidad geométrica ([2]).

En el año 1914 todavía no se identificaba a cualquier función de probabilidad con una medida pues ni siquiera estaba desarrollada la Teoría General de la Medida en espacios abstractos. En ese momento se contaba ya con la Teoría de Integración de Lebesgue y la correspondiente Teoría de la Medida en \mathbb{R}^n y eran entonces éstas las únicas medidas que al normalizarlas se consideraban probabilidades. Esto, es lo que hace Felix Hausdorff en su libro, publicado en 1914 ([57]). Ahí consideró que si A y B son dos conjuntos medibles de medida finita y $A \subset B$, entonces la medida de A dividida entre la medida de B puede considerarse como la probabilidad de que un punto que se selecciona en el conjunto B pertenezca al conjunto A . También en ese libro Hausdorff demostró el teorema de Borel sobre los números normales dentro del marco de la Teoría de la Medida.

En el libro de Hausdorff de 1914 se considera a la probabilidad como un ejemplo y una aplicación de la Teoría de la Medida. Hausdorff no identificaba a una probabilidad con una medida, pero mostró que una medida normalizada tiene todas las propiedades de una probabilidad.

El libro de Hausdorff fue durante mucho tiempo la referencia estándar para la Teoría de Conjuntos; entonces la conexión entre Probabilidad y Teoría de la Medida puede considerarse como bien establecida en la literatura matemática desde 1914.

Por otra parte, en 1913, Johann Radon había ya desarrollado una Teoría General de la Medida en \mathbb{R}^n ([89]) y en 1915, con base en el trabajo de Radon, Maurice

René Fréchet extendió la Teoría de la Medida a espacios abstractos, definiendo las funcionales aditivas ([49]). De esta manera, en ese momento se puede decir que, aunque posteriormente todavía se demostrarían algunos resultados importantes, ya se contaba con lo básico de una teoría general de la medida.

Por el lado de la Teoría de la Probabilidad, se volvió cada vez más frecuente asumir como válida ya sea la propiedad de σ -aditividad de la función de probabilidad o bien alguna de sus formas equivalentes, sobre todo en la formulación de resultados que generalizaban los teoremas límite.

En 1916, Francesco Paolo Cantelli ([20]), sin hacerlo explícito, asumió la σ -aditividad al considerar $P[|X - m| > \varepsilon]$ como una suma infinita de probabilidades.

En 1917, el mismo Cantelli ([21]), consideró que se puede asumir que la probabilidad de la intersección de una sucesión decreciente de eventos es el límite de las probabilidades de cada uno de ellos. Dice que esta propiedad responde al sentimiento de la probabilidad: “Tale assunto non pud portare ovviamente ad obbiezioni teoriche e risponde al sentimento de la probabilità, empiricamente considerata, risveglia in noi”. Concretamente, consideraba una sucesión de eventos A_1, A_2, \dots y entonces decía que se puede asumir que:

$$P\left[\bigcap_{k=1}^{\infty} A_k\right] = \lim_{n \rightarrow \infty} P\left[\bigcap_{k=1}^n A_k\right]$$

$$1 - P\left[\bigcap_{k=1}^n A_k\right] = P\left[\left(\bigcap_{k=1}^n A_k\right)^c\right] = P\left[\bigcup_{k=1}^n A_k^c\right] \leq \sum_{k=1}^n P\left[A_k^c\right]$$

Así que:

$$P\left[\bigcap_{k=1}^n A_k\right] \geq 1 - \sum_{k=1}^n P\left[A_k^c\right]$$

Por lo tanto:

$$P\left[\bigcap_{k=1}^{\infty} A_k\right] = \lim_{n \rightarrow \infty} P\left[\bigcap_{k=1}^n A_k\right] \geq 1 - \sum_{k=1}^{\infty} P\left[A_k^c\right]$$

En 1919, Guido Castelnuovo ([29]) planteó la extensión de la definición clásica de probabilidad al caso de las probabilidades en el continuo (se refería a experimentos aleatorios cuyos posibles resultados quedan representados por regiones en el plano o en el espacio tridimensional) mediante un paso al límite, con el cual quedaban extendidas las reglas de la suma y del producto al caso infinito. Consideraba sin embargo que puesta la definición en esta forma no se presta a un tratamiento matemático. Por tal motivo, la transformaba para obtener una función de densidad, asociada a cada problema particular, mediante la cual se puede obtener cualquier probabilidad como una integral de dicha función sobre una determinada región.

Todo lo anterior condujo a que para el año de 1925 algunos autores aceptaban ya a la σ -aditividad como una propiedad general de la función de probabilidad y entonces consideraban a la probabilidad como una medida. Esto queda claro en el libro de Paul Pierre Lévy, “Calcul des Probabilités”, publicado en 1925, en donde, además, se define a la probabilidad en forma axiomática ([75]). Dice Lévy:

“En el Cálculo de Probabilidades, la teoría matemática tiene como meta el establecimiento de ciertas relaciones entre las probabilidades de diferentes eventos y es independiente de las consideraciones más o menos delicadas por las cuales se determinan sus valores en las aplicaciones. Se requiere entonces darse los coeficientes de probabilidad de ciertos eventos, elegidos de tal manera que determinen perfectamente la ley de probabilidad; los otros se deducen”.

Para Lévy, dado un experimento aleatorio, los eventos, a los cuales hay que asignar probabilidades, son grupos de posibles resultados. Es decir, está ya ahí parte de la formulación moderna en la cual los eventos son subconjuntos de un conjunto Ω , llamado el espacio muestral del experimento. Finalmente, al igual que otros autores, Lévy consideraba que los dos principios fundamentales que debe satisfacer una función de probabilidad son el *Principio de las probabilidades totales* y el *Principio de la probabilidad compuesta*, pero con la diferencia de que Lévy extendía el principio de probabilidades totales al caso de una colección infinita numerable de eventos mutuamente excluyentes.

Lévy fue todavía más lejos en su formulación axiomática de la Teoría de la Probabilidad en un artículo titulado “Les lois de probabilité dans les ensembles abstraits”, publicado en el año de 1924, el cual se reprodujo al final de su libro ([74]). Dice ahí que una ley de probabilidad será naturalmente bien definida en un conjunto abstracto E si se conoce la probabilidad de todo subconjunto de E . Esta probabilidad deberá gozar de las propiedades siguientes:

- a. A dos conjuntos V_1 y V_2 sin elementos comunes y al conjunto V constituido por su unión, corresponden números α_1 , α_2 y α tales que $\alpha = \alpha_1 + \alpha_2$.
- b. Un enunciado análogo es verdadero si se considera una infinidad numerable de conjuntos V_1, V_2, \dots , sin puntos comunes dos a dos.
- c. Los valores de α son siempre positivos o nulos y al conjunto E completo corresponde un valor igual a la unidad.

Decía Lévy que, utilizando el lenguaje del Cálculo Funcional, α es una funcional aditiva en el sentido de Fréchet (es decir, una medida).

Agregaba después que en la práctica se considera una ley de probabilidad como definida sin que la probabilidad α esté definida para todos los subconjuntos de E . Cita para esto el caso en que la probabilidad de un subconjunto del intervalo $[0, 1]$ está dada por su medida de Lebesgue, en cuyo caso la probabilidad únicamente está definida para los conjuntos medibles.

Como puede verse, Lévy formuló aquí la Teoría de la Probabilidad en su forma axiomática moderna. ¿Por qué entonces se atribuye a Kolmogorov y no a Lévy esta formulación? La respuesta a esta pregunta la consideraremos más adelante.

Mientras tanto, cabe mencionar que la σ -aditividad seguía asumiéndose como válida en la formulación de los teoremas límite.

Para probar la Ley Fuerte de los Grandes Números, Aleksandr Yakovlevich Khintchine ([61]) y Andrey Nikolaevich Kolmogorov ([62]) utilizaron la propiedad de σ -subaditividad de la función de probabilidad, la cual es equivalente a la σ -aditividad. Además, Kolmogorov utilizó el hecho de que la unión numerable de eventos de probabilidad cero tiene también probabilidad cero, la cual también es consecuencia de la σ -subaditividad.

Sin embargo, la polémica sobre la propiedad de σ -aditividad de la función de probabilidad continuaba. Resalta en esta polémica una serie de artículos que publicaron Maurice René Fréchet y Bruno de Finetti en el año 1930 ([41], [42], [43], [50], [51]).

De Finetti consideraba que se llega a contradicciones cuando se admite la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos mutuamente excluyentes. Como ejemplo consideraba una variable aleatoria X la cual únicamente puede tomar valores en el conjunto infinito $\{\varepsilon_1, \varepsilon_2, \dots\}$ de tal forma que todos ellos son igualmente probables. Los eventos $[X = \varepsilon_i]$ tienen entonces probabilidad cero, pero su unión tiene probabilidad 1.

Fréchet argumentaba que él ya había señalado, en sus cursos y en una memoria que se encontraba en prensa, que efectivamente la extensión del teorema sobre las probabilidades totales al caso de una sucesión infinita de eventos no es una consecuencia inevitable de los principios generales admitidos en las bases del Cálculo de Probabilidades. Pero agregaba que de Finetti únicamente había visto una de las dos alternativas: “si sus ejemplos tienen sentido, entonces tal extensión no es posible. pero la otra alternativa es que si tal extensión es posible entonces los ejemplos no tienen sentido”. Fréchet prefería entonces asumir que los ejemplos de de Finetti no tienen sentido, en particular consideraba, con relación al mencionado ejemplo de de Finetti, que es imposible suponer que los posibles valores de X son igualmente probables. Continuaba argumentando que la misma alternativa se presenta en la teoría de la medida de Lebesgue, en donde se tiene que restringir la familia de conjuntos a

los cuales se les puede asignar una medida pues no todos los conjuntos resultan ser medibles. De la misma manera, en el ejemplo de de Fineti no es posible asignarle una probabilidad a los conjuntos $[X = \varepsilon_i]$ de tal manera que todas ellas sean iguales.

De Fineti respondió con nuevas objeciones. Se preguntaba si los eventos que se tienen que excluir de aquellos a los cuales se asigna una probabilidad no son tan interesantes como éstos últimos. Para él Fréchet únicamente evitaba formalmente la dificultad y se seguía preguntando: ¿Es admisible excluir la concepción de una infinidad de eventos mutuamente excluyentes que sean igualmente probables?

Fréchet contraargumentó que las contradicciones a que hace referencia de Fineti son familiares para todos aquellos al corriente en la teoría de la medida. En cuanto al interés que pueden tener los conjuntos no medibles responde que en realidad no se presentan en las aplicaciones. En cuanto a la necesidad de excluir algunas medidas como posibles, consideraba, por ejemplo, que se puede pensar en asignar una medida igual a 1 a toda la recta real, una medida igual a $\frac{1}{2}$ a toda semirecta, una medida igual a $\frac{1}{3}$ a todos los conjuntos formados por la unión de una sucesión infinita de intervalos de longitud λ , de tal manera que cada par de ellos esté separado por un intervalo de longitud 2λ , etc. En ese caso, toda la recta real sería la unión de una sucesión de intervalos consecutivos cuyas medidas tendrían que ser nulas, de manera que su suma no podría ser igual a 1. Por lo tanto, se debe de excluir la concepción de medidas iguales de esos intervalos o bien se deben de considerar como no medibles.

Resulta aquí claro que para Fréchet la probabilidad era siempre una medida, aún a costa de tener que excluir algunos experimentos aleatorios que pueden ser definidos formalmente, aunque también resultaba claro para él que ésta es únicamente una alternativa que se puede elegir, pero que no era aceptada por todos en ese momento. Esta posición resulta todavía más evidente al argumentar en contra de otra objeción que hace de Fineti en su segundo artículo. Decía de Fineti que no se debe eludir una dificultad de principio mediante una convención y que una vez puesta la definición de probabilidad, de una manera conforme a nuestra intuición, si esta definición permite atribuir un valor a la probabilidad de uno de los eventos clasificados como no probabilizables, no se tiene el derecho de excluir ese evento. Fréchet respondió entonces que la principal dificultad en el argumento de de Fineti reside en el hecho de que, hasta ese momento, ninguna definición de la probabilidad había obtenido una adhesión general. Agregaba que si se adopta el punto de vista axiomático, la solución es inmediata y consiste en poner como postulado el principio de las probabilidades totales en su forma completa (es decir, la propiedad de aditividad numerable). Citaba entonces que esto es lo que hace Lévy en su libro, en donde, además, justifica esta convención desde el punto de vista concreto. Más tarde, Fréchet comentaría en su libro, publicado en 1937:

“En definitiva, para adoptar el principio completo de las probabilidades totales, nos contentaremos con observar:

- 1o. Que es cómodo.
- 2o. Que cabe dentro de una teoría no contradictoria.
- 3o. Que no está en contradicción con la experiencia.”

Para entender como es que con la formulación de Kolmogorov, la cual es prácticamente la misma que la de Lévy, la adhesión a la concepción de la probabilidad como una medida fue casi unánime, debemos ver cual es la idea práctica que se encuentra detrás de la identificación de una función de probabilidad con una medida. Esta identificación permite extender el Cálculo de Probabilidades a una familia más grande de eventos. Es decir, **se trata de extender la función de probabilidad a una familia de eventos tan grande como sea posible**. Si únicamente se pide la aditividad finita, la extensión puede hacerse de manera única solamente hasta una cierta familia de eventos; mientras que si se pide la σ -aditividad la extensión puede continuarse de manera única para una familia más grande. Esta es una de las ventajas de tener la probabilidad como una medida. Sin embargo, la identificación no puede ser automática, como de hecho no lo fue, pues antes de aceptarla debe darse solución a un problema:

Mostrar que la identificación con una medida siempre es factible.

En otras palabras, **aunque ya se tenía desarrollada una Teoría de la Medida en espacios abstractos, no podía hacerse una identificación automática de una función de probabilidad con una medida mientras no se resolviera el problema de la existencia de una medida asociada a cada problema de probabilidad.**

Con relación a este problema, recordemos que el estudio de los teoremas límite había puesto en el centro de la atención de los probabilistas a las variables aleatorias. El estudio de las variables aleatorias condujo a Richard Von Mises a identificar, en el año de 1919, una ley de probabilidad con la función de distribución ([99], [100]). Esta misma identificación la hizo Lévy en su libro, en donde, además, identificaba a una función de distribución con una medida sobre \mathbb{R} y a una función de distribución conjunta con una medida sobre \mathbb{R}^n . De esta forma, **dada una sola variable aleatoria, se puede asociar a ésta una medida sobre \mathbb{R} , dado un número finito de variables aleatorias, se puede asociar a esa familia una medida sobre \mathbb{R}^n , para alguna n . Pero, ¿cómo asociarle una medida a una familia infinita de variables aleatorias?** Este problema lo atacó también Lévy en su artículo ya citado,

mostrando como se puede lograr construir una medida en una situación general; sin embargo, su método no resultó ser lo suficientemente general.

7.5. Construcción de medidas en espacios de dimensión infinita

Si bien Constantin Carathéodory, en 1914, dio un método para construir medidas en \mathbb{R}^n vía una medida exterior ([28]) y este método puede extenderse al caso de medidas en espacios abstractos ([49]), la definición de medidas en espacios de dimensión infinita no es un problema que se haya resuelto inmediatamente después del trabajo de Fréchet sobre la definición general de una medida.

Es P.J. Daniell quien entre 1918 y 1920 desarrolló una Teoría de Integración en espacios de dimensión infinita ([37], [38], [39], [40]). Daniell no se basó para esto en el resultado de Carathéodory sino que desarrolló su propio método.

Básicamente el método de Carathéodory para definir una medida consiste en partir de una medida definida sobre un álgebra de subconjuntos de un conjunto dado Ω y en extender esta medida a una σ -álgebra que contiene a los conjuntos del álgebra de la que se partió. En cambio, el método de Daniel consiste en partir de una integral definida para una cierta familia de funciones y en extender esta integral a una familia suficientemente grande de funciones. Los dos métodos son equivalentes en el sentido de que una vez teniendo una medida se puede definir una integral e inversamente, una vez teniendo una integral se puede definir una medida.

Algunos resultados parciales dentro del contexto de la Teoría de la Probabilidad se encuentran en los trabajos de Hugo Dyonizy Steinhaus ([96]) y de Norbert Wiener ([102], [103], [104], [105], [106], [107], [108]).

En 1923, Steinhaus reformuló el trabajo de Borel sobre los números normales. Para esto consideró una sucesión indefinida de ensayos de Bernoulli, en cada uno de los cuales la probabilidad de éxito es $\frac{1}{2}$, y las variables aleatorias, X_1, X_2, \dots , tales que:

$$X_j = \begin{cases} 1 & \text{si hay éxito en el ensayo } j \\ 0 & \text{si no lo hay} \end{cases}$$

El conjunto de posibles resultados del experimento aleatorio así definido consiste entonces del conjunto de sucesiones de 0's y 1's, el cual se puede poner en correspondencia, excepto por un conjunto numerable, con el intervalo $[0, 1]$.

Definió la axiomática para el juego de cara o cruz dándole a la función de probabilidad la propiedad de σ -aditividad.

Mostró entonces que comenzando por asignar probabilidades a eventos que dependen únicamente de un número finito de ensayos, las propiedades que dio a la función de probabilidad permiten definirla (extenderla) para todos los subconjuntos Lebesgue-medibles y que la medida que se obtiene es precisamente la medida de Lebesgue. Finalmente mostró que el resultado de Borel se expresa diciendo que la medida del conjunto de números normales es igual a 1.

Steinhaus consideró también el problema de la convergencia de series aleatorias de la forma $\sum_{n=1}^{\infty} \pm c_n$, en donde cada c_n es un número real y el signo de c_n se elige al azar.

Su modelo nuevamente consiste en identificar una secuencia infinita de signos como un punto del intervalo $[0, 1]$ y entonces nuevamente asumiendo que la función de probabilidad es σ -aditiva, mostró que la función de probabilidad es la medida de Lebesgue sobre los conjuntos Lebesgue-medibles. Con base en esto demostró que la probabilidad de convergencia de una serie así definida necesariamente es 0 ó 1.

En 1924, Norbert Wiener consideró también el problema de la convergencia de series aleatorias, pero su método es distinto al de Steinhaus.

Wiener trabajaba con funcionales lineales sobre espacios de funciones y seguía el método de Daniell para extender tales funcionales:

Sea Ω es el conjunto de todas las sucesiones posibles de signos. Si φ es una función definida sobre Ω cuyos valores dependen únicamente de los primeros n signos para alguna n , Wiener definió $I(\varphi)$ como el promedio de los 2^n valores que toma φ dependiendo de los primeros n signos de la sucesión. Demostró entonces que esa funcional así definida satisface las propiedades del teorema de extensión de Daniell, de manera que dicha funcional se puede extender de manera única al conjunto de todas las funciones medibles.

Con el mismo método, construyó un modelo matemático para el movimiento browniano, para lo cual definió una medida de probabilidad σ -aditiva sobre el espacio de las funciones continuas. Es este trabajo el que marcó la pauta para poder definir una medida asociada a cualquier problema de probabilidad, lo cual sería desarrollado por Kolmogorov en 1933.

7.5.1. El modelo de Kolmogorov. El modelo que formuló Kolmogorov es axiomático, lo cual se explica por el hecho de que a principios de este siglo el método axiomático había ganado un gran prestigio luego de las aportaciones de Nicolai Ivanovich Lobachevskii, Hermann Minkowski, etc., las cuales mostraban que es posible definir geometrías no euclidianas mediante diferentes sistemas axiomáticos. Aportaciones como éstas, así como la búsqueda del rigor en la ciencia, habían llevado a plantear la necesidad de la axiomatización para todas las ramas de la matemática, así como

aquellas ramas de la física en donde las matemáticas juegan un papel preponderante. Como muestra de este tipo de planteamientos basta citar el artículo de David Hilbert presentado en el décimo Congreso Internacional de Matemáticas, realizado en el año 1900 ([58]), en donde afirmó: “pienso que en cualquier lugar en donde se presenten ideas matemáticas, sea en Filosofía (Teoría del Entendimiento), sea en Geometría, sea en Física, se plantea el problema de la discusión de los principios fundamentales, base de esas ideas, y del establecimiento de un sistema simple y completo de axiomas ... Cuando se trata de plantear los principios fundamentales de una ciencia, se debe de establecer un sistema de axiomas conteniendo una descripción completa y exacta de las relaciones entre los conceptos elementales de esta ciencia”.

Con este tipo de ideas en mente, previamente al trabajo de Kolmogorov, hubo varios intentos de axiomatizar la Teoría de la Probabilidad. Entre otros se pueden citar los de Rudolf Laemmel ([65]), quien, aunque utilizaba la Teoría de la Medida únicamente de manera rudimentaria, sus axiomas incluyen la propiedad de la aditividad numerable, Ugo Broggi ([19]), quien planteaba que la probabilidad es una función no negativa definida sobre los eventos, la cual tiene la propiedad de la aditividad finita y es tal que el evento seguro tiene valor 1, afirmaba además, erróneamente, que la propiedad de la aditividad numerable es consecuencia de los axiomas, Sergi Natanovich Bernstein ([2] y [3]) y A. Lomnicki ([66]), sin embargo ninguna de ellas resultó completamente convincente.

En 1922, Borel ([17]) formuló ya la probabilidad desde un punto de vista axiomático (curso en la Facultad de Ciencias de París, publicado en 1924 como “Principes et formules classiques du Calcul des Probabilités”), sin embargo no adoptó la σ -aditividad como propiedad general de la función de probabilidad. Decía Borel:

Consideraremos eventos, adjuntándole a esa palabra la única cualidad de ser susceptibles de producirse o de no producirse.

A cada evento se le asocia un número p entre 0 y 1 de tal manera que se satisfagan dos propiedades fundamentales:

1. *Principio de las probabilidades totales:*

Dados n eventos mutuamente excluyentes, de probabilidades p_1, \dots, p_n , respectivamente, la probabilidad de que se produzca alguno de ellos es $p_1 + \dots + p_n$.

2. *Principio de la probabilidad compuesta:*

Si p_1 es la probabilidad de un evento E_1 y p_2 la probabilidad de un evento E_2 cuando E_1 se ha producido, la probabilidad de que se produzca la sucesión E_1E_2 es p_1p_2 .

Paul Lévy en su libro ([75]) retomó el punto de vista axiomático de Borel, pero consideraba ya la σ -aditividad como una propiedad general de la función de probabilidad.

Finalmente, A. N. Kolmogorov publicó su monografía ([63]), en la cual dice:

“Después de las publicaciones de las investigaciones de Lebesgue, las analogías entre medida de un conjunto y probabilidad de un evento y entre la integral de una función y la esperanza matemática de una variable aleatoria se hicieron evidentes. Pero para que la Teoría de la Probabilidad pudiera basarse en tales analogías era todavía necesario hacer las Teorías de la Medida y de la Integración independientes de los elementos geométricos los cuales estaban en el trasfondo con Lebesgue. Esto ha sido hecho por Fréchet. Mientras que una concepción de la Teoría de la Probabilidad basada sobre el punto de vista general citado antes se ha dado durante algún tiempo entre ciertos matemáticos, estaba faltando una exposición completa de todo el sistema, libre de extrañas complicaciones”.

En seguida Kolmogorov estableció los axiomas de la Teoría de la Probabilidad, primero para el caso en que únicamente entren en consideración un número finito de eventos y después para el caso general, en donde, como ya lo mencionamos. Dice que el modelo matemático de un fenómeno probabilístico está dado por una terna $(\Omega, \mathfrak{F}, P)$, en donde Ω es un conjunto, \mathfrak{F} una σ -álgebra de subconjuntos Ω y P una medida de probabilidad definida sobre \mathfrak{F} .

Partiendo de los axiomas, Kolmogorov logró entonces articular perfectamente los diferentes conceptos de la Teoría de la Probabilidad, como el de probabilidad condicional y la independencia de eventos y de variables aleatorias. Mostró además como los resultados fundamentales de la Teoría de la Probabilidad se articulan en el enfoque axiomático, exponiendo, dentro de este nuevo contexto, las leyes débil y fuerte de los grandes números.

En su monografía, Kolmogorov introdujo el concepto de Esperanza Condicional, con lo cual mostró como el enfoque axiomático basado en la Teoría de la Medida aporta a la Teoría de la Probabilidad poderosas herramientas.

Finalmente Kolmogorov dio un método general, además de simple, para construir medidas de probabilidad en espacios de dimensión infinita. Este método está basado en el resultado que él llamó el “**teorema fundamental**” y el cual establece que dada cualquier familia de variables aleatorias, partiendo de sus distribuciones finito dimensionales, es posible construir un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ de tal manera que la medida P restringida a los eventos que dependen únicamente de un número finito de las variables aleatorias dadas coincide con la determinada por la distribución finito-dimensional correspondiente.

Este resultado es de singular importancia pues muestra la consistencia de la idea de considerar a la probabilidad como una medida y demuestra la existencia de una medida asociada a cada problema de probabilidad.

Sin pretender restarle mérito al resultado de Kolmogorov, debe de mencionarse que éste no es otra cosa que una reformulación del resultado de Daniell relativo a la integración en espacios de dimensión infinita y una generalización de la idea utilizada por Wiener para construir un modelo probabilístico del movimiento browniano. La diferencia estriba en que Kolmogorov no basó su resultado en el método de Daniell para extender una funcional definida sobre un espacio de funciones sino en el método de Carathéodory para extender una medida definida sobre un álgebra de subconjuntos de un conjunto dado Ω .

La reacción que provocó la publicación del trabajo de Kolmogorov puede ilustrarse con lo que dijo Paul Lévy en el año de 1970 ([76]). Dice ahí, comentando su artículo de 1924 y publicado al final de su libro:

“El objetivo principal de mi exposición era el precisar la noción de distribución de una cierta masa igual a la unidad en un cierto espacio E . Las leyes (de probabilidad) que pueden ser definidas por una tal distribución son las que yo considero como leyes verdaderas.

Mi idea directriz era que se puede dividir el espacio E en conjuntos e_i , a los cuales se atribuyen probabilidades $\alpha_i \geq 0$ y tales que $\sum \alpha_i = 1$. Se divide en seguida cada e_i en subconjuntos $e_{i,j}$, a los cuales se atribuyen probabilidades $\alpha_{ij} \geq 0$ y de suma α_i , continuando este proceso indefinidamente.

Se llega así a una ley bien definida si cada cadena de elementos $e_i, e_{i,j}, e_{i,j,k}, \dots$, tales que cada uno es una parte del que le precede, conduce a un punto $x \in E$. Como se puede hacer la imagen de esas operaciones sobre el intervalo $[0, 1]$, se llega fácilmente al resultado siguiente: No hay leyes verdaderas a menos que el conjunto E tenga la potencia del continuo”.

Más adelante comenta sobre la parte V de su artículo:

“Creo que de cualquier manera hay un elemento positivo que se puede conservar de esa parte V. Es la idea de que la ley definida por una partición podía ser prolongada para llegar a una noción más general: una medida completamente aditiva, no negativa, definida en una familia booleana \mathcal{B} . Esta ley generalizada queda así definida por 3 elementos: el espacio E , una familia booleana \mathcal{B} de subconjuntos de E (con $E \in \mathcal{B}$) y una función m completamente aditiva, no negativa (con $m(E) = 1$ si se trata de probabilidades).

Esta tripleta es la base de la axiomática de Kolmogorov, ahora adoptada por todos los probabilistas. Cuando apareció, mi reacción fue pensar: Yo lo sabía, ¿por qué no lo dije? Me quedé durante mucho tiempo con esta idea. Pero recientemente he reflexionado que en 1924 ciertamente no me había dado cuenta de que esta idea permitía definir leyes en espacios de una potencia superior a la del continuo... Estaba yo, contrariamente a lo que había creído durante mucho tiempo, bastante lejos de haber visto bien toda la significación de la axiomática de Kolmogorov”.

Como conclusión se puede decir que **la aceptación de la probabilidad como una medida, después del trabajo de Kolmogorov, obedece en primer lugar a que Kolmogorov logró hacer una presentación clara y convincente del enfoque axiomático en la Teoría de la Probabilidad, articulando perfectamente los diferentes conceptos y los resultados fundamentales y mostrando que ese enfoque daba a la Teoría de la Probabilidad poderosas herramientas. En segundo lugar, aunque de igual o mayor importancia, el éxito obedece a que Kolmogorov logró dar un método general para construir medidas de probabilidad en espacios de dimensión infinita, mostrando así la consistencia de la idea de considerar a la probabilidad como una medida y demostrando la existencia de una medida asociada a cada problema de probabilidad.**

Referencias

- [1] Bernoulli, J., *L'Art de Conjecturer*, L.G.F. Vastel, G. Le Roy, Caen, 1801. Traducción de *Ars Conjectandi*, Basileae, 1713.
- [2] Bernstein, S. N., Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen störungen herrührendes problem, *Matematische Annalen*, 71, p. 417-439, 1911.
- [3] Bernstein, S. N., An essay on the axiomatic foundations of Probability Theory, *Proceedings of the Kharkov Mathematical Association*, Vol. 15, p. 209-274, 1917.
- [4] Bernstein, S. N., *Teoriya Veroyatnostci (Teoría de la Probabilidad)*, 1927 (4th. ed. - 1946).
- [5] Boltzmann, L., Ueber die mechanische Bedeutung des zweiten Hauptsatzes der Wärmetheorie, *Wissenschaftliche Abhandlungen* 1, 1866.
- [6] Boltzmann, L., Studien über das Gleichgewicht der lebendigen Kraft swischen bewegten materiellen Punkten, *Wissenschaftliche Abhandlungen* 1, 1868.
- [7] Boltzmann, L., *Lectures on Gas Theory (1871)*, University of California Press, Berkeley, 1964.
- [8] Boltzmann, L., Weitere Studien über das wärmeleichgewicht unter Gasmolekülen, *Wissenschaftliche Abhandlungen* 1, 1872.
- [9] Boltzmann, L., On certain questions of the theorie of gases, *Wissenschaftliche Abhandlungen* 3, 1895.
- [10] Boltzmann, L., Entgegnung auf die wärmetheoretischen Betrachtungen des Hrn. E. Zermelo, *Wissenschaftliche Abhandlungen* 3, 1896.
- [11] Boltzmann, L., *Vorlesungen über Gastheorie*, vols. 1 y 2, Barth, Leipzig, 1896-98.
- [12] Borel, F. E. J. E., Sur quelques points de la Théorie des Fonctions, *C. R. Acad. Sci.*, t. 118, p. 340-342, 1894. *Oeuvres de Émile Borel*, Tome I, Centre National de la Recherche Scientifique, p. 235-237, 1972.
- [13] Borel, F. E. J. E., Sur quelques points de la Théorie des Fonctions, Thèse doctoral, *Ann. Ec. Norm. Sup.*, 3em. série, t. 12, p. 9-55, 1895. *Oeuvres de Émile Borel*, Tome I, Centre National de la Recherche Scientifique, p. 239-285, 1972.
- [14] Borel, F. E. J. E., *Leçons sur la Théorie des Fonctions*, Gauthier-Villars, 1898.
- [15] Borel, F. E. J. E., Remarques sur certains questions de Probabilité, *Bull. Soc. Math. Fr.*, T. 32, p. 123-128, 1904. *Oeuvres de Émile Borel*, Tome II, Centre National de la Recherche Scientifique, p. 985-990, 1972.
- [16] Borel, F. E. J. E., Les probabilités dénombrables et leurs applications arithmétiques, *Rendiconti del Circolo Matematico di Palermo*, T. 27, p. 247-270, 1909. *Oeuvres de Émile Borel*, Tome II, Centre National de la Recherche Scientifique, p. 1055-1079, 1972.
- [17] Borel, F. E. J. E., *Traité du Calcul des Probabilités et de ses applications*, tome I, fascicule 1, Principes et formules classiques du Calcul des Probabilités, Gauthier-Villars, 1925. Última edición en 1947.
- [18] Borel, F. E. J. E., *Traité du Calcul des Probabilités et de ses applications*, tome II, fascicule 1, Applications a l'Arithmétique et a la Théorie des Fonctions, Gauthier-Villars, 1926.

- [19] Broggi, U., Die Axiome der Wahrscheinlichkeitsrechnung, Dissertation, Dieterich'sche Universitätsdruckerei, Göttingen, 1907.
- [20] Cantelli, F. P., Sulla legge dei grandi numeri, Mem. Acad. Lincei, Vol. 11, Série 5, p. 329-349, 1916.
- [21] Cantelli, F. P., Sulla probabilità comme limite della frequenza, Rend. Acad. Lincei, Vol. 26, p. 39-45, 1917.
- [22] Cantelli, F. P., Su due applicazioni di un teorema di G. Boole alla Statistica Matematica, Accademia dei Lincei Roma, Classe di Scienze Fisiche, Matematiche e Naturali, Rendiconti, 26 (5), p. 295-302, 1917.
- [23] Cantor, G. F. L. P., Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 1, Math. Ann., 15, p. 1-7, 1879.
- [24] Cantor, G. F. L. P., Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 2, Math. Ann., 17, p. 355-358, 1880.
- [25] Cantor, G. F. L. P., Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 3, Math. Ann., 20, p. 113-121, 1882.
- [26] Cantor, G. F. L. P., Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 4, Math. Ann., 21, p. 51-58 y 545-591, 1883.
- [27] Cantor, G. F. L. P., Ueber unendliche, lineare Punktmannichfaltigkeiten, Pt. 5, Math. Ann., 23, p. 453-488, 1884.
- [28] Caratheodory, C., Über das lineare Mass von Punktmengen, Nachrichten von der Königlichen Gesellschaft der Wiss zu Göttingen, p. 404-426, 1914.
- [29] Castnuovo, G., Calcolo delle Probabilità, Nicola Zanichelli, Bologna, 1919 (2a. ed. - 1925).
- [30] Cauchy, A. L., Résumé des leçons données a l'École Royale Polytechnique sur le Calcul Infinitésimal, Imprimerie Royale, 1823.
- [31] Clausius, R. J. E., Über die Art der Bewegung, welche wir Wärme nennen, Annalen der Physik und Chemie, 100, 1857.
- [32] Clausius, R. J. E., Ueber die mittlere Länge der Wege, welche bei der Molecularbewegung gasförmiger Körper von den einzelnen Molecülen zurückgelegt werden; nebst einigen anderen Bemerkungen über die mechanische Wärmetheorie, Annalen der Physik, 105, 1858.
- [33] Clausius, R. J. E., On the second fundamental theorem of the mechanical theory of heat, Philosophical Magazine, 35, 1868.
- [34] Chebyshev, P. L., Des valeurs moyennes, Matematiceskii Sbornik, 127, p. 1-9, 1867, también publicado en Liouville's Journal de Mathématiques Pures et Appliquées, 88, p.177-184, 1867.
- [35] Chebyshev, P. L., Démonstration élémentaire d'une proposition générale de la théorie des probabilités.
- [36] Chebyshev, P. L., Sur deux théorèmes relatifs aux probabilités.
- [37] Daniell, P. J., A general form of integral, Annals of Mathematics, Vol. 19, 1918.
- [38] Daniell, P. J., Functions of limited variation in an infinite number of dimensions, Annals of Mathematics, serie II, Vol. 21, p. 30-38, 1920.
- [39] Daniell, P. J., Further properties of the general integral, Annals of Mathematics, Serie II, Vol. 21, p. 203-220, 1920.
- [40] Daniell, P. J., Integrals in an infinite number of dimensions, Annals of Mathematics.
- [41] de Finetti, B., Sui passaggi al limite nel Calcolo delle Probabilità, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 155-166, 1930.
- [42] de Finetti, B., A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 901-905, 1930.

- [43] de Finetti, B., Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 1063-1069, 1930.
- [44] de Moivre, A., The doctrine of chances, A. Millar, London, 1718 (third edition - 1756). Reimpreso por Chelsea, New York, 1967.
- [45] Dirichlet, J. P. G. L., 1829.
- [46] Drach, J.
- [47] du Bois-Reymond, P. D. G., Über die Integration der trigonometrischen Reihe, Math. Ann., 22, p. 260-268, 1883.
- [48] du Bois Reymond, P., Über die Integration der Reihen, berlin Ak. Sber., p. 359-371, 1886.
- [49] Fréchet, M. R., Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, Bull. Soc. Mat. de France, 43, 1915.
- [50] Fréchet, M. R., Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements, (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 899-900, Milano, 1930.
- [51] Fréchet, M. R., Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements (seconde note), (Reale) Istituto Lombardo de Science e Lettere, Rendiconti, Vol. 63, p. 1059-1062, 1930.
- [52] Gibbs, J. W., Elementary Principles in Statistical Mechanics, 1902 Reimpreso por Dover, New York, 1962.
- [53] Hankel, H., Untersuchungen über die unendlich oft oszillierenden und unstetigen Functionen, University of Tübingen, 1870, reproducido en Math. Ann., 20, 1882.
- [54] Harnack, A., Die elemente der Differential und Integralrechnung, B. G. Teubner, Leipzig, 1881.
- [55] Harnack, A., Lehrbuch der Differential und Integralrechnung, 2 Vols., B. G. Teubner, Leipzig, 1884-1885.
- [56] Harnack, A., Über den Inhalt von Punktmengen, Math. Ann. 25, p. 241-250, 1885.
- [57] Hausdorff, F., Grundzüge der Mengenlehre, Chelsea Publishing Company, 1914.
- [58] D. Hilbert, Sur les problèmes futures des Mathématiques, Comptes Rendus du Deuxième Congrès International des mathématiciens, Paris, p. 58-114, 1900.
- [59] Huygens, C., Du calcul dans les jeux de hasard, Oeuvres Complètes de Christiaan Huygens, Vol. XIV, Martinus Nijhoff, 1920. Traducción de De Ratiociniis in Aleae Ludo, 1657.
- [60] Jordan, M. E. C., Cours d'Analyse de l'École Polytechnique, 3 Vols., Gauthier-Villars, 1882-1887. (Second edition, 1893-1896; Third edition, 1909).
- [61] Khintchine, A. Ya., Sur la loi forte des grands nombres, C. R. Ac. Sc. Paris, Vol. 186, p. 285-287, 1928.
- [62] Kolmogorov, A. N., Sur la loi forte des grands nombres, C. R. Ac. Sc. Paris, Vol. 191, p. 910-912, 1930.
- [63] Kolmogorov, A. N., Foundations of the Theory of Probability, Chelsea, 1950. Traducción de Grundbegriffe der Wahrscheinlichkeitsrechnung, Erg Mat. 2, No. 3, 1933.
- [64] Krönig, A., Grundzüge, einer Theorie der Gase, Annalen der Physic und Chemie, 99, 1856.
- [65] Laemmel, R., Untersuchungen über die Ermittlung der Wahrscheinlichkeiten, Dissertation, Zurich, 1904.
- [66] Lomnicki, A., Nouveaux fondements du Calcul des Probabilités, Fundamenta Mathematicae, t. 4, p. 34-71, 1923.
- [67] Laplace, P. S., Mémoire sur la probabilité des causes par les evenements, Mémoires de l'Académie Royale des Sciences de Paris (Savants étrangers), Tome VI, p. 621, 1774. Oeuvres complètes de Laplace, Tome huitième, Gauthier-Villars, 1891.
- [68] Laplace, P. S., Mémoire sur les Probabilités, Mémoires de l'Académie Royale des Sciences de Paris, 1778. Oeuvres complètes de Laplace, Tome neuvième, Gauthier-Villars, 1893.

- [69] Laplace, P. S., *Théorie Analytique des Probabilités* (1812), Livre I. Calcul des fonctions génératrices, Troisième édition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.
- [70] Laplace, P. S., *Théorie Analytique des Probabilités* (1812), Livre II. Théorie générale des probabilités, Troisième édition, Courcier, Paris, 1820. Oeuvres complètes de Laplace, Tome septième, Gauthier-Villars, 1886.
- [71] Laplace, P. S., *Essai philosophique sur les Probabilités* (1814), Gauthier-Villars, 1921.
- [72] Lebesgue, H. L., Intégrale, longueur, aire, Thèse doctoral, Ann. Math. Pur. Appl., 7 (3), p. 231-359, 1902.
- [73] Lebesgue, H. L., *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthier-Villars, 1904.
- [74] Lévy, P. P., Les lois de probabilité dans les ensembles abstraits, Revue de Métaphysique et Morale, 1924. Reproducido en *Calcul des Probabilités*, Gauthier Villars, 1925.
- [75] Lévy, P. P., *Calcul des Probabilités*, Gauthier Villars, Paris, 1925.
- [76] Lévy, P. P., *Premiers travaux sur le Calcul des Probabilités*, 1970, Oeuvres, Vol. III, 1976.
- [77] Lipschitz, R. O. S., De explicatione per series trigonometricas instituenda functionum unius variabilis arbitrariarum et praecipue earum, quae per variabilis spatium finitum valorum maximorum et minimorum numerum habent infinitum, disquisitio, Crelle, JI. Math., 63, 1864, traducción al francés en *Acta Math.*, 36, 1912.
- [78] Lyapunov, A. M., Sur une proposition de la Théorie des Probabilités, *Izv. Akad. Nauk.*, Ser. 5, 13, p. 359-386, 1900.
- [79] Lyapunov, A. M., Nouvelle forme du théorème sur la limite des probabilités, *Notes Acad. Sci. Phys. Math. Sect.*, Ser. 8, 2, p. 1-24, 1901.
- [80] Markov, A. A., The law of large numbers and the method of least squares, *Izd. Fiz. Mat. Ob. va Pri Kazan*, Ser. 2, 8, p. 110-128, 1898.
- [81] Markov, A. A., Sur les racines de l'équation $\frac{e^{x^2} \delta^m e^{-x^2}}{\delta x^m} = 0$, *Izv. Akad. Nauk.*, Ser. 5, 9, p. 435-446, 1898.
- [82] Markov, A. A., Extensión de la ley de los grandes números a variables dependientes, *Notices (Izvestiya) of the Physical Mathematical Society at Kazan University*, Ser. 2, 15 (no.4), p. 155-156, 1907.
- [83] Markov, A. A., Teorema del Límite Central para variables aleatorias dependientes, 1908, 1910, 1911, 1912.
- [84] Markov, A. A., *Ischislenie Veroyatnostei (El Cálculo de Probabilidades)*, Moscow, 1913 (Cuarta edición, 1924).
- [85] Maxwell, J. C., On the dynamical theory of gases, 1867, *Scientific Papers of James Clerk Maxwell*, vol. 2.
- [86] Maxwell, J. C., Does the progress of Physical Science tend to give any advantage to the opinion of Necessity (or Determinism) over that of Contingency of Events and the freedom of the Will?, 1873, *The life of James Clerk Maxwell*, 1882.
- [87] Peano, G., *Applicazione geometriche del Calcolo Infinitesimale*, Torino, 1887.
- [88] Poincaré, J. H., *Calcul des Probabilités*, Gauthier-Villars, Paris, 1896.
- [89] Radon, J., *Theorie u. Anwendungen der absolut additiven Mengenfunktionen*, Sitzber der Math Naturwiss, Klasse der Kais, Akademie der Wiss, Wien, 1913.
- [90] Bhaskara Rao, K.P.S. and Bhaskara Rao, M., *Theory of Charges (A study of finitely additive measure)*, Academic Press, 1983.
- [91] Riemann, G. F. B., Sur la possibilité de représenter une fonction par une série trigonométrique, *Mémoires de la Société Royale des Sciences de Göttingue*, t. XIII, 1867, traducción al francés reproducida en *Oeuvres Mathématiques de Riemann*, A. Blanchard, Paris, 1968.

- [92] Sierpinski, W., Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre, Bull. Soc. Math. France, t. 45, p. 125-132, 1917.
- [93] Smith, H. J. S., On the integration of discontinuous functions, London Math. Soc. Proc., 6, 1875.
- [94] Stolz, O., Über einen zu einer unendlichen Punktmenge gehörigen Grenzwert, Math. Ann., 23, p. 152-156, 1884.
- [95] Stolz, O., Grundzüge der Differential und Integralrechnung, 3 Vols., B. G. Teubner, Leipzig, 1893-99.
- [96] Steinhaus, H. D., Les probabilités dénombrables et leur rapport à la Théorie de la Mesure, Fundamenta Mathematicae, t. 4, p. 286-310, 1923.
- [97] Volterra, V., Alcune osservazioni sulle funzioni punteggiate discontinue, Giorn. Mat., 19, p. 76-86, 1881.
- [98] Volterra, V., Sui principii del Calcolo Integrale, Giorn. Mat., 19, p. 333-372, 1881.
- [99] Von Mises, R., Grundlagen der Wahrscheinlichkeitsrechnung, Math. Zeitsch., Vol. 5, p. 52-99, 1919.
- [100] Von Mises, R., Mathematical Theory of Probability and Statistics, 1919.
- [101] Von Plato, J., Creating modern Probability, Cambridge University press, 1994.
- [102] Wiener, N., The mean of a functional of arbitrary elements, Ann. of Math., (2) 22, p. 66-72, 1920.
- [103] Wiener, N., The average of an analytic functional, Proc. Nat. Acad. Sci. U.S.A., Vol. 7, No. 9, p. 253-260, 1921.
- [104] Wiener, N., The average of an analytic functional and the Brownian Movement, Proc. Nat. Acad. Sci. U.S.A., Vol. 7, No. 10, p. 294-298, 1921.
- [105] Wiener, N., Differential space, J. Math. and Physics, 2, p. 131-174, 1923.
- [106] Wiener, N., Note on the series $\sum (\pm \frac{1}{n})$, Bull. Acad. Polon. Ser. A, 13, p. 83-90, 1923.
- [107] Wiener, N., Un problème de probabilités dénombrables, Bull. Soc. Math. France 11, p. 569-578, 1924.
- [108] Wiener, N., The average value of a functional, Proc. London Math. Soc., 22, p. 454-467, 1924.

Respuestas a los ejercicios

CAPÍTULO 1

1.1. a) sí lo es; b) no lo es.

$$1.2. f_{X,Y}(x, y) = \begin{cases} \frac{1}{4}(xy - \frac{1}{4}x^2 - y^2 + x + 2y + 1) & \text{si } 0 \leq y < -\frac{1}{2}x + 1 \leq 1 \\ \frac{1}{4}(-\frac{1}{2}x^2 - 2y^2 + 2x + 4y) & \text{si } 0 < -\frac{1}{2}x + 1 \leq y < 1 \\ \frac{1}{4}(-\frac{1}{2}x^2 + 2x + 2) & \text{si } 0 \leq x < 2, 1 \leq y \\ \frac{1}{4}(-2y^2 + 4y + 2) & \text{si } 0 \leq y < 1, 2 \leq x \\ 1 & \text{si } 2 \leq x, 1 \leq y \end{cases}$$

$$1.4. f_{X,Y}(x, y) = \begin{cases} \frac{10!}{x!y!(10-x-y)!} p^x q^y & \text{si } x, y \in \{0, \dots, 10\}, x + y \leq 10 \\ 0 & \text{en otro caso} \end{cases}$$

1.5. a) $X_1 + X_2$ tiene distribución binomial con parámetros n y $p = p_1 + p_2$.

b) Dado que $X_1 + X_2 = z$, X_2 tiene distribución binomial con parámetros z y $p = \frac{p_2}{p_1 + p_2}$.

$$1.6. f_{X_1, \dots, X_r}(x_1, \dots, x_r) = \begin{cases} \frac{(2r)!}{x_1! \dots x_r!} \left(\frac{1}{r}\right)^{2r} & \text{si } \sum_{k=1}^r x_k = 2r, x_k \in \{0, \dots, 2r\} \\ 0 & \text{en otro caso} \end{cases}$$

$$1.8. f_{X_1, X_2, X_3}(n_1, n_2, n_3) = \begin{cases} \frac{\binom{m_1}{n_1} \binom{m_2}{n_2} \binom{m_3}{n_3}}{\binom{m_1 + m_2 + m_3}{n}} & \text{si } n_1 + n_2 + n_3 = n, n_k \in \{0, \dots, m_k\} \\ 0 & \text{en otro caso} \end{cases}$$

1.9. a) $\frac{c}{2}N^2(N^2 + 1)$; b) $\frac{c}{6}N^2(N^4 - 1)$; c) $1 - \frac{c}{6}N^2(N^2 + 1)(N^2 + 2)$

1.10. $1 - e^{-1}$

$$1.11. F_{X, X^2}(x, y) = \begin{cases} 0 & \text{si } y < 0 \\ F_X(x) - F_X(-\sqrt{y}) & \text{si } y \geq x^2 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & \text{si } 0 \leq y \leq x^2 \end{cases}$$

No existe una función de densidad conjunta.

1.12. No existe.

$$1.13. f_X(x) = \begin{cases} c(N^2 - k)x & \text{si } x \in \{k^2 + 1, \dots, (k+1)^2\}, k \in \{0, \dots, N-1\} \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{cy^2(y^2+1)}{2} & \text{si } y \in \{1, \dots, N\} \\ \frac{cN^2(N^2+1)}{2} & \text{si } y \in \{N+1, \dots, N^2\} \\ 0 & \text{en otro caso} \end{cases}$$

$$1.14. a) F_{X,Y}(x, y) = \begin{cases} 0 & \text{si } x \leq 0 \text{ ó } y \leq 0 \\ 2xy & \text{si } 0 \leq x \leq 1, 0 \leq y \leq 1-x \\ 1 - (1-y)^2 - (1-x)^2 & \text{si } 0 \leq x \leq 1, 1-x \leq y \leq 1 \\ 1 - (1-y)^2 & \text{si } x \geq 1, 0 \leq y \leq 1 \\ 1 - (1-x)^2 & \text{si } y \geq 1, 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1, y \geq 1 \end{cases}$$

$$b) f_X(x) = \begin{cases} 2(1-x) & \text{si } x \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} 2(1-y) & \text{si } y \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$1.15. f_{X,Y}(x, y) = \begin{cases} \frac{1}{4} & \text{si } (x, y) \text{ pertenece al interior del rombo} \\ 0 & \text{en otro caso} \end{cases}$$

$$f_X(x) = \begin{cases} \frac{1}{4}(2+x) & \text{si } x \in (-2, 0] \\ \frac{1}{4}(2-x) & \text{si } x \in (0, 2) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} 1+y & \text{si } x \in (-1, 0] \\ 1-y & \text{si } x \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$1.16. f_X(x) = \begin{cases} \frac{1}{4}x^3 & \text{si } x \in (0, 2) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{4}y(4-y^2) & \text{si } y \in (0, 2) \\ 0 & \text{en otro caso} \end{cases}$$

$$1.17. a) \frac{1}{2}$$

b) X tiene distribución exponencial con parámetro λ .

Y tiene distribución gama con parámetros 2 y λ .

$$1.18. f_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

$$= \begin{cases} (1-x_1) \prod_{k=2}^n \left[\frac{2}{3} |x_k - x_{k-1}| + \frac{1}{3} (1 - (x_k - x_{k-1})) \right] & \text{si } x_1, \dots, x_n \in \{0, 1\} \\ 0 & \text{en otro caso} \end{cases}$$

1.19. *Sí lo son.*

1.20. $\frac{1}{n+m+1}$

1.21. a) $\frac{p(1-p)^3}{1-(1-p)^4}$; b) $\frac{2p(1-p)}{2-p}$; c) $p(2-p) + \frac{p(1-p)^2}{2-p}$

1.22. $1 - e^{-\lambda p}$

1.23. $\frac{1}{2^{2n}} \binom{2n}{n}$

1.24. a) $\frac{N+2}{2(N+1)}$; b) $\frac{1}{N+1}$

1.25. $\frac{N-1}{4N}$

1.26. a) $\frac{1}{N}(1-p)^3 [1 - (1-p)^N] + \frac{1}{N} [1 - (1-p)^{N-1}]$

b) $\frac{1-p}{Np} [1 - (1-p)^N]$

1.27. $\frac{1}{4}$

1.28. a) $p(1-p)^2 e^{-\lambda p}$; b) $e^{-\lambda p}$

1.29. 0.281874

1.30. $\frac{1}{3}$

1.31. $1 - (\lambda + 1)e^{-\lambda}$

1.32. $\frac{1}{3}$

1.33. $\frac{4}{9}$

1.34. $\frac{4\lambda_1\lambda_2}{(2\lambda_1+\lambda_3)(2\lambda_2+\lambda_3)}$

1.35. a) $\frac{1}{2}$; b) $\frac{2}{3}$; c) $\frac{17}{24}$

1.36. $1 - \frac{1}{27}e^{-\frac{1}{2}\lambda}(6\lambda + 20) - \frac{1}{27}e^{-3\lambda}(9\lambda + 7)$

1.37. $e^{-\frac{1}{2}} - e^{-1}$

1.38. $\frac{1}{4}(1 - 2\lambda e^{-2\lambda} - e^{-2\lambda})$

1.39. $\frac{17}{18}$

1.40. $f_{X^2, Y^2}(u, v)$

$$= \begin{cases} \frac{1}{4\sqrt{uv}} [f(\sqrt{u}, \sqrt{v}) + f(\sqrt{u}, -\sqrt{v}) + f(-\sqrt{u}, \sqrt{v}) + f(-\sqrt{u}, -\sqrt{v})] & \text{si } u > 0, v > 0 \\ 0 & \text{en otro caso} \end{cases}$$

1.41. a) $\frac{1}{6}$; b) T tiene distribución uniforme en el intervalo $(0, 15)$.

1.42. a) $\frac{1}{36} (5 + 6 \ln 2)$; b) 0

CAPÍTULO 2

2.1. a) $f_{X+Y}(z) = \begin{cases} \frac{\lfloor \frac{z}{2} \rfloor + 1}{28} & \text{si } z \in \{0, \dots, 6\} \\ \frac{\lfloor \frac{z}{2} \rfloor - z + 7}{28} & \text{si } z \in \{7, 8, \dots, 12\} \\ 0 & \text{en otro caso} \end{cases}$

b) $f_{Y-X}(z) = \begin{cases} \frac{7-z}{28} & \text{si } z \in \{0, \dots, 6\} \\ 0 & \text{en otro caso} \end{cases}$

2.2. a) $f_{X+Y}(z) = \begin{cases} \frac{1}{380}(z-2) & \text{si } z \in \{4, 6, \dots, 20\} \\ \frac{1}{380}(z-1) & \text{si } z \in \{3, 5, \dots, 19\} \\ \frac{1}{380}(40-z) & \text{si } z \in \{22, 24, \dots, 38\} \\ \frac{1}{380}(41-z) & \text{si } z \in \{21, 23, \dots, 39\} \\ 0 & \text{en otro caso} \end{cases}$

b) $f_{Y-X}(z) = \begin{cases} \frac{20-z}{190} & \text{si } z \in \{1, \dots, 19\} \\ 0 & \text{en otro caso} \end{cases}$

2.3. $\frac{1-(1-p)^{x+1}}{2+(1-p)^x(p-2)}$

2.4. a) $f_{X+Y}(z) = \begin{cases} \frac{z-1}{N^2} & \text{si } z \in \{2, \dots, N\} \\ \frac{2N-z+1}{N^2} & \text{si } z \in \{N+1, \dots, 2N\} \\ 0 & \text{en otro caso} \end{cases}$

b) $f_{\min(X,Y)}(z) = \begin{cases} \frac{2(N-z)+1}{N^2} & \text{si } z \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$

c) $f_{\max(X,Y)}(z) = \begin{cases} \frac{2z-1}{N^2} & \text{si } z \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$

d) $f_{Y-X}(z) = \begin{cases} \frac{N-|z|}{N^2} & \text{si } z \in \{1-N, \dots, N-1\} \\ 0 & \text{en otro caso} \end{cases}$

e) $f_{|Y-X|}(z) = \begin{cases} \frac{1}{N} & \text{si } z = 0 \\ \frac{2(N-z)}{N^2} & \text{si } z \in \{1, \dots, N-1\} \\ 0 & \text{en otro caso} \end{cases}$

2.5. $f_Z(z) = \begin{cases} \frac{4z^3}{n^2(n+1)^2} & \text{si } z \in \{1, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$

2.6. $f_X(x) = \begin{cases} \frac{2x-1}{144} & \text{si } x \in \{1, \dots, 12\} \\ 0 & \text{en otro caso} \end{cases}$

$$2.7. f_Z(z) = \begin{cases} \frac{(1-p)^z}{N} [1 + p(N - z)] & \text{si } z \in \{1, \dots, N\} \\ p & \text{si } z = 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.8. f_{X+Y}(z) = \begin{cases} \frac{1}{4} & \text{si } z = 0 \\ \frac{1}{2N} + \frac{z-1}{4N^2} & \text{si } z \in \{1, \dots, N\} \\ \frac{1}{2N} - \frac{z-1}{4N^2} & \text{si } z \in \{N+1, \dots, 2N\} \\ 0 & \text{en otro caso} \end{cases}$$

2.11. 0.277124

2.12. a) $X + Y$ tiene distribución uniforme en el conjunto $\{1, \dots, 2N\}$.

b) $X - Y$ tiene distribución uniforme en el conjunto $\{1 - N, \dots, N\}$.

$$2.13. f_{X+Y}(z) = \begin{cases} \frac{2z}{3N^2(N+1)^2} (z^2 - 1) & \text{si } z \in \{2, \dots, N\} \\ \frac{2(2N+1-z)}{3N^2(N+1)^2} (z^2 + z + 2zN - 2N^2 - 2N) & \text{si } z \in \{N+1, \dots, 2N\} \\ 0 & \text{en otro caso} \end{cases}$$

$$2.14. \binom{z}{y} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^y \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{z-y}$$

Es decir, dado que $X + Y = z$, X tiene distribución binomial con parámetros $n = z$ y $p = \frac{\lambda_2}{\lambda_1 + \lambda_2}$.

$$2.15. \frac{\binom{m}{y} \binom{n}{z-y}}{\binom{n+m}{z}}$$

Es decir, dado que $X + Y = z$, Y tiene distribución hipergeométrica con parámetros m , n y z .

$$2.16. \frac{1}{6}$$

$$2.17. \frac{n}{(n+1)(2n+1)}$$

$$2.18. \frac{2n}{(n+1)(2n+1)}$$

$$2.21. a) f_{2X+Y}(z) = \begin{cases} \frac{z}{2} & \text{si } z \in (0, 1) \\ \frac{1}{2} & \text{si } z \in [1, 2) \\ \frac{3-z}{2} & \text{si } z \in [2, 3) \\ 0 & \text{en otro caso} \end{cases}$$

$$b) f_{3X-Y}(z) = \begin{cases} \frac{1}{3}(1+z) & \text{si } -1 < z < 0 \\ \frac{1}{3} & \text{si } 0 \leq z < 2 \\ \frac{1}{3}(3-z) & \text{si } 2 \leq z < 3 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.22. f_{4Y-3X}(z) = \begin{cases} \frac{1}{7}\lambda e^{\frac{1}{3}\lambda z} & \text{si } z < 0 \\ \frac{1}{7}\lambda e^{-\frac{1}{4}\lambda z} & \text{si } z \geq 0 \end{cases}$$

$$2.23. 0.945431$$

$$2.25. 1245$$

$$2.26. 0.195586$$

2.27. $X^2 + Y^2$ tiene distribución exponencial con parámetro $\lambda = \frac{1}{2}$.

$$2.28. f_{X+Y}(z) = \begin{cases} 1 - e^{-\lambda z} & \text{si } 0 < z < 1 \\ e^{-\lambda z}(e^\lambda - 1) & \text{si } z \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.29. f_{|Y-X|}(z) = \begin{cases} \frac{2(b-a-z)}{(b-a)^2} & \text{si } 0 \leq z < b-a \\ 0 & \text{en otro caso} \end{cases}$$

$$2.30. a) f_{X+Y}(z) = \begin{cases} \frac{1}{4}ze^{-\frac{1}{2}z} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$b) f_{\max(X,Y)}(z) = \begin{cases} \frac{1}{6}z^3e^{-z} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.32. c = \frac{1}{4}; f_{Y-X}(z) = \begin{cases} \frac{1}{4}(3-z) & \text{si } z \in (0, 2) \\ 0 & \text{en otro caso} \end{cases}$$

$$2.33. c = \frac{2}{11}$$

$$f_X(x) = \begin{cases} \frac{3c}{2}(4-x^2) & \text{si } x \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{c}{2}(7-2y) & \text{si } y \in (0, 2) \\ \frac{3c}{2}(y-3)^2 & \text{si } y \in [2, 3) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_{2Y-3X}(z) = \begin{cases} \frac{c}{72}(z+3)(33-z) & \text{si } z \in (-3, 0) \\ \frac{c}{8}(11-2z) & \text{si } z \in [0, 3) \\ \frac{c}{8}(4-z)(3z-4) & \text{si } z \in [3, 4) \\ 0 & \text{en otro caso} \end{cases}$$

$$2.34. f_W(w) = \frac{1}{|b|} \int_{-\infty}^{\infty} f(x, \frac{w-ax}{b}) dx$$

$$2.35. f_U(u) = \begin{cases} \frac{1}{8}(u+3) & \text{si } u \in (-3, -1) \\ \frac{1}{4} & \text{si } u \in [-1, 1) \\ \frac{1}{8}(3-u) & \text{si } u \in [1, 3) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_V(v) = \begin{cases} \frac{1}{8}(v+3) & \text{si } v \in (-3, -1) \\ \frac{1}{4} & \text{si } v \in [-1, 1) \\ \frac{1}{8}(3-v) & \text{si } v \in [1, 3) \\ 0 & \text{en otro caso} \end{cases}$$

2.36. $f_V(v) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{v}{x}) dx$

2.37. $f_{XY}(z) = \begin{cases} \frac{1}{3}e^{-\sqrt{-z}} & \text{si } z < 0 \\ \frac{1}{6}e^{-\sqrt{z}} & \text{si } z > 0 \\ 0 & \text{si } z = 0 \end{cases}$

2.38. $f_{XY}(z) = \begin{cases} -\ln z & \text{si } z \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$

2.39. $f_{XY}(z) = \begin{cases} -\frac{1}{2} \ln(-z) & \text{si } z \in (-1, 0) \\ -\frac{1}{2} \ln z & \text{si } z \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$

$P[-\frac{1}{4} < XY < \frac{1}{2}] = \frac{3}{8} + \frac{1}{2} \ln 2$

2.40. Z tiene distribución normal estándar.

2.41. $f_{\frac{Y}{X}}(z) = \begin{cases} \frac{1}{z^2} & \text{si } z > 1 \\ 0 & \text{en otro caso} \end{cases}$

2.42. $f_Z(z) = \begin{cases} \frac{1}{3z^2} & \text{si } z > 1 \\ \frac{1}{3z^2} & \text{si } z < -1 \\ 0 & \text{en otro caso} \end{cases}$

$P[-3 < \frac{Y}{X} < 2] = \frac{11}{18}$

2.43. $f_{\frac{Y}{X}}(z) = \frac{1}{2\lambda} \left(1 - \frac{\lambda}{|z|} e^{-\frac{\lambda}{|z|}} - e^{-\frac{\lambda}{|z|}} \right)$ para $z \neq 0$.

2.44. $f_{\frac{Y}{X}}(z) = \frac{\alpha}{2\lambda} F_U\left(\frac{1}{|z|}\right)$ para $z \neq 0$, en donde F_U es la función de distribución de una variable aleatoria con distribución gama de parámetros $\alpha + 1$ y λ .

2.45. $f_Z(z) = \begin{cases} \frac{1}{\beta(\alpha_1, \alpha_2)} \frac{z^{\alpha_2-1}}{(1+z)^{\alpha_1+\alpha_2}} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$

2.46. $f_Z(z) = \frac{1}{\pi \left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} z^2 \right)}$

2.47. $f_U(u) = \begin{cases} \frac{1}{u^2} \ln u & \text{si } u > 1 \\ 0 & \text{en otro caso} \end{cases}$

V tiene distribución uniforme en el intervalo $(0, 1)$.

2.52. $\min(X, Y)$ tiene distribución exponencial con parámetro 2λ .

$$f_{\max(X, Y)}(z) = \begin{cases} 2\lambda e^{-\lambda z}(1 - e^{-\lambda z}) & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.53. \ a) \ f_{X+Y+Z}(w) = \begin{cases} \frac{w^2}{2} & \text{si } w \in (0, 1) \\ 1 - \frac{(w-1)^2}{2} - \frac{(2-w)^2}{2} & \text{si } w \in [1, 2) \\ \frac{(3-w)^2}{2} & \text{si } w \in [2, 3) \\ 0 & \text{en otro caso} \end{cases}$$

$$b) \ f_{X+Y-Z}(w) = \begin{cases} \frac{1}{2}(w+1)^2 & \text{si } w \in (-1, 0) \\ 1 - \frac{1}{2}w^2 - \frac{1}{2}(w-1)^2 & \text{si } w \in [0, 1) \\ \frac{1}{2}(2-w)^2 & \text{si } w \in [1, 2) \\ 0 & \text{en otro caso} \end{cases}$$

2.54. a) Y tiene distribución exponencial con parámetro $n\lambda$.

$$b) \ f_Z(z) = \begin{cases} \lambda n e^{-\lambda z} (1 - e^{-\lambda z})^{n-1} & \text{si } z > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.55. \ \frac{5}{8}$$

2.57. N_t tiene distribución Poisson con parámetro λt .

2.58. f alcanza su valor máximo en $x = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$.

$$2.59. \ f_{X+Y, Y-X}(u, v) = \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda u} & \text{si } -u < v < u \\ 0 & \text{en otro caso} \end{cases}$$

$$P[X + Y \leq 1, Y - X \geq 0] = \frac{1}{2} [1 - e^{-\lambda(\lambda + 1)}]$$

$$2.60. \ f_{X+Y, Y-X}(u, v) = \begin{cases} \frac{1}{2}\lambda^2 e^{-\frac{1}{2}\lambda(u+v)} & \text{si } 0 < v < u \\ 0 & \text{en otro caso} \end{cases}$$

$$P[X + Y \leq 1, Y - X \geq 0] = 1 + e^{-\lambda} - 2e^{-\frac{1}{2}\lambda}$$

$$2.61. \ f_{U, V}(u, v) = \begin{cases} \frac{1}{2}(u^2 - v^2) & \text{si } v - 2 < u < -v < 0 \text{ ó } 0 < v < u < -v + 2 \\ 0 & \text{en otro caso} \end{cases}$$

$$P[U < 1, V > \frac{1}{2}] = \frac{17}{192}$$

$$2.66. \ f_{U, V}(u, v) = \begin{cases} \lambda^2 e^{\lambda v} & \text{si } 0 < u < 2v \\ 0 & \text{en otro caso} \end{cases}$$

$$2.68. F_{U,V}(u_0, v_0) = \begin{cases} \frac{1}{B(\alpha, \alpha)} \int_{1-v_0}^{u_0} [u(1-u)]^{\alpha-1} du & \text{si } 0 \leq 1-v_0 \leq u_0 \leq 1 \\ \frac{1}{B(\alpha, \alpha)} \int_{1-v_0}^1 [u(1-u)]^{\alpha-1} du & \text{si } 0 \leq 1-v_0 \leq 1 \text{ y } u_0 \geq 1 \\ \frac{1}{B(\alpha, \alpha)} \int_0^{u_0} [u(1-u)]^{\alpha-1} du & \text{si } 1-v_0 \leq 0 \text{ y } 0 \leq u_0 \leq 1 \\ 1 & \text{si } 1-v_0 \leq 0 \text{ y } u_0 \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

No existe una función de densidad conjunta.

$$2.69. f_{X+Y, \frac{X}{X+Y}}(u, v) = \begin{cases} u & \text{si } uv < 1, u(1-v) < 1, u \in (0, 2), v \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

U y V no son independientes.

$$2.70. f_{U,V}(u, v) = \begin{cases} \frac{u}{(1-v)^2} \lambda^2 e^{-\frac{\lambda u}{1-v}} & \text{si } u > 0 \text{ y } v \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

U y V no son independientes.

$$2.71. f_{U,V}(u, v) = \begin{cases} \frac{1}{2\pi\sqrt{u(v-u)}} e^{-\frac{v}{2}} & \text{si } 0 < u < v \\ 0 & \text{en otro caso} \end{cases}$$

U y V no son independientes.

$$2.72. f_{R,\Theta}(r, \theta) = \begin{cases} \frac{r}{\pi} & \text{si } r \in [0, 1), \theta \in [0, 2\pi) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_R(r) = \begin{cases} 2r & \text{si } r \in [0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{si } \theta \in [0, 2\pi) \\ 0 & \text{en otro caso} \end{cases}$$

R y Θ son independientes.

$$2.73. f_{R,\Phi}(r, \theta) = \begin{cases} \frac{r}{\pi} & \text{si } r \in [0, 1), \theta \in [0, 2\pi) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_R(r) = \begin{cases} 2r & \text{si } r \in [0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$f_\Phi(\theta) = \begin{cases} \frac{1}{2\pi} & \text{si } \theta \in [0, 2\pi) \\ 0 & \text{en otro caso} \end{cases}$$

R y Φ son independientes.

$$2.74. f_{U,V}(u, v) = \begin{cases} \frac{u}{v^2} & \text{si } u < v, u \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$P[U > \frac{1}{2}, V < 2] = \frac{5}{16}$$

$$2.75. f_{U,V}(u, v) = \begin{cases} \frac{1}{4u} & \text{si } -u < v < u, u \in (0, 1) \\ -\frac{1}{4u} & \text{si } u < v < -u, u \in (-1, 0) \\ 0 & \text{en otro caso} \end{cases}$$

$$P[U < \frac{1}{2}, V < \frac{1}{2}] = \frac{5}{8} + \frac{1}{8} \ln 2$$

$$2.76. a) f_{U,V}(u, v) = \begin{cases} \frac{2}{uv^2} & \text{si } 1 < u < v < u^2 \\ 0 & \text{en otro caso} \end{cases}$$

$$b) P[2 < U < 4, V < 9] = \frac{13}{36} - \frac{4}{9} \ln 2 + \frac{2}{9} \ln 3$$

$$c) f_U(u) = \begin{cases} \frac{2(u-1)}{u^3} & \text{si } u > 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_V(v) = \begin{cases} \frac{1}{v^2} \ln v & \text{si } v > 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.77. a) f_{U,V}(u, v) = \begin{cases} \frac{1}{2u^2v} & \text{si } 0 < \frac{1}{u} \leq v \leq u \\ 0 & \text{en otro caso} \end{cases}$$

$$b) P[U > 2, V < 3] = \frac{1}{3} + \frac{1}{2} \ln 2$$

$$c) f_U(u) = \begin{cases} \frac{1}{u^2} \ln u & \text{si } u \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$f_V(v) = \begin{cases} \frac{1}{2} & \text{si } v \in (0, 1) \\ \frac{1}{2v^2} & \text{si } v \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$2.78. b) P[X + Y > Z] = \frac{1}{2}$$

$$2.81. f_{T_1, T_2}(t_1, t_2) = \begin{cases} 4\lambda^2 e^{-2\lambda(t_1+t_2)} & \text{si } t_1, t_2 \in (0, \infty) \\ 0 & \text{en otro caso} \end{cases}$$

T_1 y T_2 son independientes y ambas tienen distribución exponencial con parámetro 2λ .

$$2.82. f_{T_1, T_2}(t_1, t_2) = \begin{cases} 4(1-t_2) - 2t_1 & \text{si } t_1, t_2 \in (0, 1) \text{ y } t_1 + t_2 < 1 \\ 0 & \text{si } t_1 + t_2 \geq 1 \end{cases}$$

T_1 y T_2 no son independientes.

2.83. $f_{P_{t_1}, \dots, P_{t_n}}(x_1, \dots, x_n)$

$$= \begin{cases} \lambda^{x_n} e^{-\lambda t_n} \frac{t_1^{x_1} (t_2 - t_1)^{x_2 - x_1} \dots (t_n - t_{n-1})^{x_n - x_{n-1}}}{x_1! (x_2 - x_1)! \dots (x_n - x_{n-1})!} & \text{si } x_1, \dots, x_n \in \{0, 1, \dots\} \text{ y } x_1 \leq x_2 \leq \dots \leq x_n \\ 0 & \text{en otro caso} \end{cases}$$

2.84. $f_{W_{t_1}, \dots, W_{t_n}}(y_1, \dots, y_n)$

$$= \frac{1}{\sqrt{(2\pi)^n t_1 (t_2 - t_1) \dots (t_n - t_{n-1})}} \exp \left\{ -\frac{1}{2} \left[\frac{1}{t_1} y_1^2 + \frac{1}{t_2 - t_1} (y_2 - y_1)^2 + \dots + \frac{1}{t_n - t_{n-1}} (y_n - y_{n-1})^2 \right] \right\}$$

2.85. $\frac{1}{2^{n+1}}$

2.86. $e^{-\frac{n(n-1)\lambda}{2}}$

2.87. $\frac{k}{n+1}$

2.88. a) $\frac{8}{27}$; b) $\frac{91}{216}$; c) $60(1 - \frac{1}{\sqrt[3]{2}})$ minutos.

2.89. $\frac{5}{36}$

2.90. $\frac{31}{32}$

2.91. $P[X_{(1)} > \frac{1}{2}, X_{(2)} < 2] = e^{-\frac{3}{2}} - 3e^{-\frac{9}{2}} + 2e^{-6}$

$P[X_{(2)} < 1, X_{(3)} > 1] = 3e^{-1} - 6e^{-2} + 3e^{-3}$

2.92. $P[X_{(1)} > -\frac{1}{2}, X_{(2)} < \frac{1}{2}] = \frac{5}{16}$

$P[X_{(2)} < \frac{1}{2}, X_{(3)} > \frac{1}{2}] = \frac{27}{64}$

2.93. $\frac{7}{27}$

2.94. $E[Z] = \begin{cases} 0 & \text{si } k \in \{2, 3, \dots\} \\ \infty & \text{si } k = 1 \end{cases}$

$Var(Z) = \begin{cases} \frac{k}{k-2} & \text{si } k \in \{3, 4, \dots\} \\ \infty & \text{si } k = 2 \end{cases}$

2.95. $E[Z] = \begin{cases} \frac{m}{m-2} & \text{si } m \in \{3, 4, \dots\} \\ \infty & \text{si } m \in \{1, 2\} \end{cases}$

$Var(Z) = \begin{cases} \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)} & \text{si } m \in \{5, 6, \dots\} \\ \infty & \text{si } m \in \{3, 4\} \end{cases}$

2.96. $E[Z] = \frac{\alpha}{\alpha+\beta}$; $Var(Z) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

2.97. $\frac{287}{40}$

2.98. $\frac{553}{40}$

2.99. a) $\frac{1}{6N} (N+1)(2N+1)$; b) $\frac{1}{3N} (N^2-1)$

2.101. $\frac{1}{4}$

2.102. $\sum_{k=0}^{\infty} \frac{1}{k!} = e$

2.103. $\sum_{k=0}^{\infty} \frac{1}{k!} = e$

2.104. $\sum_{k=1}^{\infty} \frac{1}{k+1} = \infty$

2.105. a) $\frac{n}{n+1}$; b) $\frac{1}{n+1}$

2.106. 1

2.107. $\sqrt{\frac{2}{\pi^2-6}}$

CAPÍTULO 3

3.1. $E[X] = 1$, $E[Y] = -2$, $Var(X) = 13$, $Var(Y) = 5$ y $\rho = \frac{2}{\sqrt{65}}$

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{61}} \exp \left\{ -\frac{1}{2} \left[\frac{5}{61}(x-1)^2 + \frac{13}{61}(y+2)^2 - \frac{4}{61}(x-1)(y+2) \right] \right\}$$

3.2. b) $f(x,y) = \frac{\sqrt{31}}{4\pi} \exp \left\{ -\frac{1}{2} \left[2 \left(x + \frac{22}{31} \right)^2 + 3 \left(x + \frac{22}{31} \right) \left(y - \frac{19}{31} \right) + 5 \left(y - \frac{19}{31} \right)^2 \right] \right\}$

c) $\mu_X = -\frac{22}{31}$, $\mu_Y = \frac{19}{31}$, $\sigma_X^2 = \frac{20}{31}$, $\sigma_Y^2 = \frac{8}{31}$ y $\rho = -\frac{3}{20}\sqrt{10}$

3.3. $U = \frac{\sqrt{6}}{4}X - \frac{\sqrt{3}}{12}Y - \frac{\sqrt{6}}{8} - \frac{\sqrt{3}}{24}$; $V = \frac{1}{\sqrt{6}}Y + \frac{1}{2\sqrt{6}}$

3.4. d) *La distribución conjunta de X, Y no es normal bivariada.*

3.5. $\frac{1}{4} + \frac{1}{2\pi} \arcsen \rho$

3.6. $f_Y(y) = \frac{1}{\sigma\sqrt{6\pi}} e^{-\frac{1}{6\sigma^2}(y-\mu_1-\mu_2)^2}$

3.7. $f_Y(y) = \frac{1}{\sqrt{42\pi}} e^{-\frac{1}{42}(y-6)^2}$

3.8. $f_{U,V}(u,v) = \frac{1}{5\pi\sqrt{3}} \exp \left\{ -\frac{14}{25} \left[\frac{u^2}{7} + \frac{v^2}{3} - \frac{3}{21}uv \right] \right\}$

3.9. a) $A^{-1} = \begin{pmatrix} 3 & -1 & -1 \\ 1 & -\frac{1}{2} & 0 \\ -6 & \frac{5}{2} & 2 \end{pmatrix}$

b) $B^{-1} = \frac{1}{128} \begin{pmatrix} 25 & -62 & 25 & 21 \\ -75 & -70 & -75 & 65 \\ 80 & -160 & -240 & 80 \\ 35 & -10 & 35 & 55 \end{pmatrix}$

c) C No es invertible.

$$d) D^{-1} = \begin{pmatrix} 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & -1 & \frac{1}{2} \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} & 1 \end{pmatrix}$$

3.10. a) no lo es; b) sí lo es; c) sí lo es; d) no lo es

3.11. a) sí lo es; b) no lo es; c) sí lo es

$$3.12. P = \begin{pmatrix} \frac{1}{3}\sqrt{3} & -\frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{3} & \frac{1}{2}\sqrt{2} & -\frac{1}{6}\sqrt{6} \\ \frac{1}{3}\sqrt{3} & 0 & \frac{1}{3}\sqrt{6} \end{pmatrix}$$

3.13. a) Sí lo es.

$$B = \sqrt{\frac{4+\sqrt{5}}{110}} \begin{pmatrix} \sqrt{33-11\sqrt{5}} + \sqrt{23-3\sqrt{5}} & 4\sqrt{2} - \sqrt{10} - \sqrt{22} \\ 4\sqrt{2} - \sqrt{10} - \sqrt{22} & \sqrt{33+11\sqrt{5}} + \sqrt{103-45\sqrt{5}} \end{pmatrix}$$

b) No lo es.

c) Sí lo es.

$$B \approx \begin{pmatrix} 0.8621 & -0.3673 & 0.2633 & -0.2294 \\ -0.3673 & 1.1254 & -0.5966 & 0.4926 \\ 0.2633 & -0.5966 & 1.3547 & -0.8599 \\ -0.2294 & 0.4926 & -0.8599 & 1.7220 \end{pmatrix}$$

$$3.15. a) P = \begin{pmatrix} \frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{3} & \frac{1}{2} \\ -\frac{1}{2}\sqrt{2} & \frac{1}{6}\sqrt{6} & \frac{1}{6}\sqrt{3} & \frac{1}{2} \\ 0 & -\frac{1}{3}\sqrt{6} & \frac{1}{6}\sqrt{3} & \frac{1}{2} \\ 0 & 0 & -\frac{1}{2}\sqrt{3} & \frac{1}{2} \end{pmatrix}$$

$$b) B = \frac{1}{8} \begin{pmatrix} 3\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} \\ -\sqrt{3} + \sqrt{7} & 3\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} \\ -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & 3\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} \\ -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & -\sqrt{3} + \sqrt{7} & 3\sqrt{3} + \sqrt{7} \end{pmatrix}$$

c) $A = B$

$$3.16. a) A = \begin{pmatrix} \frac{1}{3}\sqrt{3} & \frac{5}{6} & -\frac{17}{30}\sqrt{5} \\ 0 & 1 & -\frac{1}{5}\sqrt{5} \\ 0 & 0 & \frac{3}{5}\sqrt{5} \end{pmatrix}$$

$$b) A = \begin{pmatrix} \frac{1}{2}\sqrt{10} & -3\sqrt{5} & -\frac{1}{6}\sqrt{15} & -\frac{37}{6} \\ 0 & \sqrt{5} & 0 & 2 \\ 0 & 0 & \frac{1}{3}\sqrt{15} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$c) A = \begin{pmatrix} \frac{1}{5}\sqrt{5} & -\frac{2}{5} & \frac{2}{5}\sqrt{3} & \frac{4}{5}\sqrt{2} & \frac{8}{15}\sqrt{2} \\ 0 & 1 & 0 & -\sqrt{2} & -\frac{2}{3}\sqrt{2} \\ 0 & 0 & \sqrt{3} & \sqrt{2} & \frac{2}{3}\sqrt{2} \\ 0 & 0 & 0 & \sqrt{2} & \frac{2}{3}\sqrt{2} \\ 0 & 0 & 0 & 0 & \sqrt{2} \end{pmatrix}$$

$$3.17. A = \begin{pmatrix} \frac{\sqrt{14}}{4} & \frac{\sqrt{10}}{20} & \frac{\sqrt{15}}{20} & \frac{1}{4} \\ 0 & \frac{3\sqrt{10}}{10} & \frac{\sqrt{15}}{20} & \frac{1}{4} \\ 0 & 0 & \frac{\sqrt{15}}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$3.18. a) \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}; b) U \text{ y } V \text{ son independientes.}$$

$$c) \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}; d) X \text{ y } Y \text{ no son independientes.}$$

$$3.19. a) f_{U,V}(\bar{u}) = \frac{1}{6\pi} \exp \left\{ -\frac{1}{2} (Q\bar{u}) \cdot \bar{u} \right\}, \text{ en donde } Q = \begin{pmatrix} \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{9} \end{pmatrix}$$

$$b) C = \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}$$

3.20. 0.755969

3.21. U tiene distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 4$.

$\frac{1}{4}V$ tiene distribución χ^2 con 3 grados de libertad.

3.22. $\sqrt{2}Z$ tiene distribución t con 2 grados de libertad.

$$3.23. f_{Y_1, Y_2, Y_3, Y_4}(y_1, y_2, y_3, y_4) = \frac{1}{4\pi^2} e^{-\frac{1}{2}(y_1^2 + y_2^2 + y_3^2 + y_4^2)}$$

Z tiene distribución uniforme en el intervalo $(0, 1)$.

3.28. U tiene distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 2$.

CAPÍTULO 4

$$4.1. a) \frac{8N+5}{6}; b) \frac{N^2(2N+1)-5}{2N^2-3}$$

4.2. a) $\frac{5}{3\lambda}$; b) $\frac{2}{3\lambda}$

4.3. a) $\frac{11}{6}$; b) 0.55416

4.4. a) $\frac{1}{6}$; b) $\frac{2}{3}$

4.5. a) $-\frac{3}{10}$; b) 0

4.6. $\frac{1}{2Y+N+1} \left[\frac{1}{3} (N+1)(2N+1) + (N+1)Y \right]$

4.7. a) $\frac{1}{3}(Y+1)$; b) $\frac{1}{3}(X+1+2N)$

4.8. a) $\frac{1}{2}Y$; b) $\frac{1}{2}X + \frac{1}{2}(N+1)$

4.9. a) $\frac{1}{2}(N+1) - \frac{1}{2}(Y-X)$; b) $\frac{1}{2}(N+1) + \frac{1}{2}(Y-X)$

4.10. $\min(X, Y) + \frac{(1-p)}{p(2-p)}$

4.11. $\frac{\max(X, Y)[3\max(X, Y)-1]}{2[2\max(X, Y)-1]}$

4.12. a) $Y - \frac{1}{2N}Y(Y-1)$; b) $\frac{1}{2}(N+1) + \frac{1}{2N}Y(Y-1)$

4.20. a) 0; b) $6e^{-2|X|}$

4.21. a) $\frac{1}{2}Y$; b) $\frac{X^3-3X^2+6X-6}{X^2-2X+2}I_{(-\infty,0)}(X) + \frac{2(X^2+2X+2)}{1+X}I_{(0,\infty)}(X)$

4.22. $\frac{1}{6} \frac{12Y^3-30Y^2-36Y-13}{2Y-7} I_{(0,2)}(Y) + \frac{1}{18}(35Y^2 - 34Y + 21)I_{[2,3)}(Y)$

4.23. $\frac{6}{\alpha^2}Y + Y^3$

4.24. $\frac{1}{Y+1}e^{-Y}$

4.25. a) $Y^2 + \frac{2}{\lambda}Y + \frac{2}{\lambda^2}$; b) $Y + \frac{1}{\lambda}$

4.26. a) 0; b) 0

4.27. a) $\frac{2}{3}I_{(0,1)}\left(\frac{X}{Y}\right) + \frac{2}{3} \frac{Y}{X} I_{[1,\infty)}\left(\frac{X}{Y}\right)$; b) $\frac{1}{2Y}$

4.28. $Y - X + \frac{2}{\lambda}$

4.29. $X + \frac{1}{\lambda}$

4.30. $\frac{\alpha+\beta}{\lambda}$

4.31. $\frac{1}{6}(X+Y)^2$

4.32. $E[X | U] = UI_{[0,t)}(U) + (1+t)I_{\{t\}}(U)$

$E[X | V] = VI_{(t,\infty)}(V) + \left(1 - \frac{te^{-t}}{1-e^{-t}}\right) I_{\{t\}}(V)$

4.33. 0

$$4.34. f_{X|Y}(x | y) = \begin{cases} \frac{2(x+y)}{N(N+1+2y)} & \text{si } x \in \{1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

$$4.35. a) f_{X|Y}(x | y) = \begin{cases} \frac{2(y-x)}{y^2-y} & \text{si } x \in \{1, \dots, y-1\} \\ 0 & \text{en otro caso} \end{cases}$$

$$b) f_{Y|X}(y | x) = \begin{cases} \frac{2(y-x)}{(N-x)(N+1-x)} & \text{si } y \in \{x+1, \dots, N\} \\ 0 & \text{en otro caso} \end{cases}$$

4.36. a) Dado que $Y = y$, X tiene distribución uniforme en el conjunto $\{1, \dots, y-1\}$.

b) Dado que $X = x$, Y tiene distribución uniforme en el conjunto $\{x+1, \dots, N\}$.

$$4.37. a) f_{\max(X,Y)|Y}(u | y) = \begin{cases} \frac{y}{N} & \text{si } u = y \\ \frac{1}{N} & \text{si } u \in \{y+1, \dots, N\} \\ 0 & \text{si } u \in \{1, \dots, y-1\} \end{cases}$$

$$b) f_{\min(X,Y)|X}(v | x) = \begin{cases} \frac{N-x+1}{N} & \text{si } v = x \\ \frac{1}{N} & \text{si } v \in \{1, \dots, x-1\} \\ 0 & \text{si } v \in \{x+1, \dots, N\} \end{cases}$$

4.38. Dado que $X_2 = n$, X_1 tiene distribución uniforme en el conjunto $\{1, \dots, n-1\}$.

4.39. Para $z \in \mathbb{N}$, dado que $X+Y = z$, X tiene distribución binomial con parámetros $n = z$ y $p = \frac{\lambda_2}{\lambda_1 + \lambda_2}$. Para $z = 0$, la distribución de X , dado que $X+Y = z$, está concentrada en $x = 0$.

4.40. a) Dado que $N_i = s$, la distribución de N_j es binomial con parámetros $n - s$ y $p = \frac{p_j}{1-p_i}$.

$$b) \text{Cov}(N_i, N_j) = -np_i p_j$$

4.41. Si $0 < x < 1$, dado que $X = x$, Y tiene distribución uniforme en el intervalo $(0, 2x)$, mientras que si $1 \leq x < 2$, dado que $X = x$, Y tiene distribución uniforme en el intervalo $(0, 4 - 2x)$.

$$4.42. a) f_{X_{(3)}|X_{(1)}}(x_3 | x_1) = \begin{cases} \frac{2(x_3-x_1)}{(1-x_1)^2} & \text{si } 0 < x_1 < x_3 < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$b) \frac{X_{(1)}+2}{3}$$

$$4.43. a) f_{U,V}(u, v) = \frac{1}{5\pi\sqrt{3}} \exp \left\{ -\frac{14}{25} \left[\frac{u^2}{7} + \frac{v^2}{3} - \frac{3}{21} uv \right] \right\}; b) \frac{V}{2}$$

4.44. 0.6816

$$4.45. a) \frac{2}{3}(X+Y-Z); b) \frac{14}{3} + \frac{4}{9}(X+Y-Z)^2$$

4.46. Si $z \in (0, 1)$, dado que $X + Y = z$, X tiene distribución uniforme en el intervalo $(0, z)$ y, si $z \in [1, 2)$, dado que $X + Y = z$, X tiene distribución uniforme en el intervalo $(z - 1, 1)$.

4.47. Dado que $X + Y = z$, X tiene distribución normal con parámetros $\mu = \frac{z}{2}$ y $\sigma^2 = \frac{1}{2}$.

4.48. a) $\frac{2}{3}Y$

$$b) f_{X|Y-X}(x | z) = \begin{cases} \frac{6x(x+z)}{(2+z)(1-z)^2} & \text{si } 0 < x < 1 - z, z \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

c) 0.652038

4.49. a) $\frac{2}{3}Y$; b) $\frac{2}{3}[1 - (Y - X)]$

c) $\frac{175}{296}$; d) $2 - \frac{2}{Y} + \frac{2e^{-Y}}{Y}$

4.50. a) Si $z \leq 0$, dado que $Y - X = z$, Y tiene distribución exponencial con parámetro 2λ y si $z > 0$, dado que $Y - X = z$, $Y - z$ tiene distribución exponencial con parámetro 2λ .

b) $\frac{1}{2\lambda} + (Y - X)I_{(0, \infty)}(Y - X)$

4.51. Dado que $Y - X = z$, la distribución de X es uniforme, en el intervalo $(-z, 1)$ si $z < 0$ y en el intervalo $(0, 1 - z)$ si $z \geq 0$.

Dado que $Y - X = z$, la distribución de Y es uniforme, en el intervalo $(0, 1 + z)$ si $z < 0$ y en el intervalo $(z, 1)$ si $z \geq 0$.

4.52. Dado que $Y - X = z$, X tiene distribución normal con media $-\frac{z}{2}$ y varianza $\frac{3}{4}$.

$$4.53. a) f_{X, Y-4X}(x, z) = \frac{1}{\pi\sqrt{15}} \exp \left\{ -\frac{32}{15} \left[x^2 + \frac{z^2}{16} + \frac{7xz}{16} \right] \right\}$$

b) Dado que $Y - 4X = z$, X tiene distribución normal con media $-\frac{7}{32}z$ y varianza $\frac{15}{64}$.

4.54. a) Dado que $Y - X = v$, $X + Y$ tiene distribución uniforme en el intervalo $(|v|, 2 - |v|)$.

b) 1

$$4.55. P[\text{máx}(X, Y) \leq z | Y = y] = \begin{cases} 0 & \text{si } z < y \\ z & \text{si } y \leq z < 1 \\ 1 & \text{si } z \geq 1 \end{cases}$$

$$E[\text{máx}(X, Y) | Y] = \frac{1}{2}(1 + Y^2)$$

$$4.56. \ a) \ P[\min(X, Y) \leq z \mid Y = y] = \begin{cases} 0 & \text{si } z \leq 0 \\ 1 - e^{-\lambda z} & \text{si } 0 < z < y \\ 1 & \text{si } z \geq y \end{cases}$$

$$P[\max(X, Y) \leq z \mid Y = y] = \begin{cases} 0 & \text{si } z < y \\ 1 - e^{-\lambda z} & \text{si } z \geq y \end{cases}$$

$$b) \ E[\min(X, Y) \mid Y] = \frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda Y}$$

$$E[\max(X, Y) \mid Y] = Y + \frac{1}{\lambda}e^{-\lambda Y}$$

4.59. a) X tiene distribución Poisson con parámetro λp .

b) Dado que $X = x$, $Y - x$ tiene distribución Poisson con parámetro $\lambda(1 - p)$.

$$4.60. \ a) \ E(Y) = \mu; \text{Var}(Y) = a^2 + \sigma^2; \text{Cov}(X, Y) = \sigma^2$$

$$c) \ f_Y(y) = \frac{1}{\sqrt{2\pi\sqrt{a^2 + \sigma^2}}} \exp\left\{-\frac{1}{2(a^2 + \sigma^2)}(y - \mu)^2\right\}$$

$$d) \ E[X \mid Y = y] = \frac{a^2}{a^2 + \sigma^2}\mu + \frac{\sigma^2}{a^2 + \sigma^2}y$$

Por lo tanto, dado que $Y = y$, el mejor estimador de X , en el sentido de la media cuadrática, es $\frac{a^2}{a^2 + \sigma^2}\mu + \frac{\sigma^2}{a^2 + \sigma^2}y$.

$$4.61. \ E(Y) = \alpha + \frac{1}{2}\beta; \text{Var}(Y) = \sigma^2 + \frac{1}{12}\beta^2$$

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y - \alpha - \beta x)^2\right\}$$

4.62. 6 horas.

$$4.63. \ \frac{N(N+1)}{2}$$

$$4.64. \ \frac{1-p^r}{p^r(1-p)}$$

$$4.65. \ \frac{1}{2^n}$$

$$4.66. \ \frac{1}{2}$$

$$4.67. \ \frac{r+s}{3} + \frac{2r-s}{3} \left(\frac{r+s}{r+s+3}\right)^n$$

$$4.68. \ [1 + (2p - 1)^2]^n x$$

$$4.69. \ a + r - a \left(1 - \frac{1}{a+r}\right)^n$$

$$4.70. \ a) \ \frac{1}{2}; \ b) \ \frac{1}{3}$$

4.71. $E[Z] = 12.5$, en donde Z es el total de trabajadores afectados por algún accidente en una semana.

Si el número de trabajadores afectados en un accidente particular no fuera independiente del número de accidentes, entonces no se podría asegurar la misma respuesta.

$$4.72. f_Y(y) = \begin{cases} -\ln y & \text{si } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

4.73. a) X_T tiene distribución geométrica de parámetro $p = \frac{\nu}{\lambda + \nu}$.

$$b) P[X_T = k] = \begin{cases} \frac{1}{\lambda a} \left[1 - e^{-\lambda a} \sum_{j=1}^k \frac{1}{j!} (\lambda a)^j \right] & \text{si } k \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

$$4.74. a) P[X = x] = \begin{cases} \frac{\binom{\alpha+x-1}{x} \binom{\beta+n-x-1}{n-x}}{\binom{\alpha+\beta+n-1}{n}} & \text{si } x \in \{0, \dots, n\} \\ 0 & \text{en otro caso} \end{cases}$$

$$b) E[X] = \frac{n\alpha}{\alpha+\beta}; \text{Var}(X) = \frac{n\alpha\beta}{(\alpha+\beta)^2} \frac{\alpha+\beta+n}{\alpha+\beta+1}$$

c) Dado que $X = x$, Y tiene distribución beta con parámetros $\alpha + x$ y $\beta + n - x$.

4.75. a) ∞

b) Dado que $X = x$, Y tiene distribución beta con parámetros 2 y $x + 1$.

$$c) X^2 + \frac{4X}{X+3} + \frac{6}{(X+3)(X+4)}$$

$$4.76. a) P[X = x] = \begin{cases} \frac{B(\alpha+1, \beta+x)}{B(\alpha, \beta)} & \text{si } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

$$b) \frac{\beta}{\alpha-1}$$

c) Dado que $X = x$, Y tiene distribución beta con parámetros $\alpha + 1$ y $\beta + x$.

$$d) \frac{\alpha+1}{\alpha+\beta+X+1}$$

$$4.77. a) \frac{r\beta}{\alpha-1}$$

b) Dado que $X = x$, Y tiene distribución beta de parámetros $\alpha + r$ y $\beta + x$.

$$c) X^2 + \frac{6X}{X+4} + \frac{12}{(X+4)(X+5)}$$

$$4.78. a) P[X = k] = \begin{cases} \frac{1}{k+1} - \frac{1}{k+2} & \text{si } k \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

$$b) E[X] = \infty$$

c) Dado que $X = x$, Y tiene distribución beta con parámetros 2 y $x + 1$.

$$d) \frac{2}{X+3}$$

$$4.79. a) P[X = x] = \begin{cases} \frac{\Gamma(\alpha+x)}{x!\Gamma(\alpha)} \left(\frac{\lambda}{\lambda+1}\right)^\alpha \left(\frac{1}{\lambda+1}\right)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{en otro caso} \end{cases}$$

En el caso particular en que α es un entero positivo, X tiene distribución binomial negativa con parámetros α y $p = \frac{\lambda}{\lambda+1}$.

$$b) E[X] = \frac{\alpha}{\lambda}$$

c) Dado que $X = x$, Y tiene distribución gama con parámetros $\alpha + x$ y $\lambda + 1$.

$$d) \frac{\alpha+X}{\lambda+1}$$

$$4.80. b) \frac{\lambda+1}{(\lambda+2)^3}; c) \frac{3(\lambda+1)^2}{(\lambda+2)^4}$$

CAPÍTULO 5

5.8. Es mayor o igual a $\frac{3}{5}$.

$$5.10. \approx \Phi(16.531 \ln a - 18.127)$$

en donde Φ es la función de distribución normal estándar.

$$5.11. a) P\left[\sum_{i=1}^{20} X_i > 15\right] \leq \frac{20}{15}$$

$$b) P\left[\sum_{i=1}^{20} X_i > 15\right] \approx 0.86822$$

$$5.12. \approx 0.58014$$

$$5.13. F_Z(z) \approx \Phi\left(\frac{z - \frac{2n}{\lambda^2}}{\frac{2}{\lambda^2} \sqrt{5n}}\right)$$

5.14. Por lo menos 175.

$$5.15. a) \approx 0.99972; b) \approx 0.97672$$

$$5.16. \approx 0.16769$$

$$5.17. \approx 0.6816$$

5.18. Utilizando el teorema del límite central, $P[|X - 10| \leq 2] \approx 0.8427$.

Utilizando la desigualdad de Chebyshev, $P[|X - 10| \leq 2] \geq \frac{1}{2}$.

$$5.19. 24$$

Índice

- Álgebra de subconjuntos, 298
- Aditividad finita
 - propiedad de la, 298
- Aditividad numerable
 - propiedad de la, 298
- aleatorio
 - vector, 10
- Bayes
 - fórmula de, 200
 - teorema de, 203
- Bayes, T., 200
- Bernoulli
 - teorema de, 234
- Bernoulli, J., 219, 287, 292, 293, 295, 305
- Bernstein, S. N., 332, 340
- Boltzmann, L., 301
- Borel, F. E. J. E., 247, 307, 326, 340
- Borel-Cantelli
 - lema de, 229, 316, 317
- Broggi, U., 340
- Brown, R., 81
- Cálculo de Probabilidades
 - clásico, 300
 - surgimiento del, 271
- Cantelli, F. P., 333
- Cantor, G. F. L. P., 324
- Carathéodory C., 338
- Cardano, G., 272, 275
- Castelnuovo, G., 333
- Cauchy, A. L., 319
- Cauchy-Schwarz
 - desigualdad de, 69
- Chebyshev, P. L., 220, 234, 259, 300
- Clausius, R. J. E., 301
- Coefficiente de correlación, 71
- Conjunto
 - de Cantor, 324
 - de contenido cero, 325
 - de medida cero, 326
 - de primera especie, 324
 - denso en ninguna parte, 320
 - Jordan medible, 326
 - medible, 330
- Contenido
 - exterior, 326
 - interior, 326
 - teoría del, 326
- Convergencia
 - casi segura, 225
 - de series aleatorias, 263
 - diferentes tipos de, 220
 - en distribución, 224
 - en probabilidad, 220
 - relación entre modos de, 226
- Convolución, 41
- Covarianza, 68
- Covarianzas
 - matriz de, 71
- d'Imola, B., 273
- Daniell, P. J., 338
- De Finetti, B., 335
- de Fournival, R., 273
- de Méré, A. G., 277
- de Moivre, A., 201, 220, 240, 287, 306
- Dedekind, J. W. R., 321
- Dirichlet, J. P. G. L., 320
- Distribución
 - beta, 54
 - condicional, 171

- caso absolutamente continuo, 174
 - caso discreto, 151
 - caso mixto, 193
- de Polya, 214
- F, 50
- marginal, 8, 20
- multinomial, 13, 14
- normal bivariada, 85
- normal estándar
 - tabla, 369
- normal multivariada, 125
- t, 48
- Distribución conjunta, 3
 - de variables aleatorias independientes, 24
 - función de, 6
 - propiedades, 8, 10
- Drach, J. J., 328
- du Bois-Reymond, P. D. G., 324
- Esperanza, 66
 - condicional, 149
 - caso absolutamente continuo, 164
 - caso discreto, 151
 - dada la ocurrencia de un evento, 149
 - definición general, 156
 - propiedades, 160
 - interpretación, 240
- Estadísticos de orden, 61
- Feller, W., 220
- Fermat, P., 219, 271
- Feynman, R. P., 3
- Finitamente aditiva
 - función, 298
- Formas cuadráticas, 88, 101
 - definidas positivas, 88, 101
- Fréchet, M. R., 333, 335
- Freud, S., v
- Función de densidad
 - conjunta, 11
 - de vectores aleatorios absolutamente continuos, 15
 - de vectores aleatorios discretos, 11, 13
 - marginal, 21
- Funciones de vectores aleatorios
 - absolutamente continuos
 - distribución de, 40
 - discretos
 - distribución de, 37
 - distribuciones conjuntas, 51
 - esperanza de, 66
- Funciones generadoras, 231
- Galileo, G., 273
- Gauss, J. C. F., 85
- Gauss-Jordan
 - método de, 98
- Gibbs, J. W., 301
- Gram-Schmidt
 - proceso de ortogonalización de, 110, 118
- Hadamard, J. S., 149
- Hankel, H., 323
- Harnack, A., 325, 326
- Hausdorff, F., 332
- Heisenberg, W., 3
- Helmert
 - transformación de, 140
- Heráclito, 219
- Hilbert, D., 301, 340
- Huygens, C., 219, 271, 278, 304
- Independencia de variables aleatorias
 - criterio para la, 25
- Integrabilidad
 - criterio de Lebesgue, 329
 - primer criterio de Riemann, 322
 - reformulación del segundo criterio de Riemann, 326
 - segundo criterio de Riemann, 322
- Integral
 - de Cauchy, 319
 - de Riemann, 321
- Jordan, M. E. C., 326
- Khintchine, A. K., 220, 236, 239, 335
- Kolmogorov
 - desigualdad de, 250
- Kolmogorov, A. N., 220, 246, 297, 335, 341
- Krönig, A., 301
- Lévy, P. P., 219, 220, 300, 334, 341
- Laemmel, R., 340
- Lagrange
 - metodo de, 101
- Laplace, P. S., 300
- Lebesgue, H. L., 329

- Ley débil de los grandes números, 234
 - resultado de Bernoulli-Chebyshev, 235
 - resultado de Khintchine, 239
 - resultado de Poisson-Markov, 235
- Ley fuerte de los grandes números, 246
 - resultado de Borel, 247
 - resultado de Kolmogorov, 251, 256
 - resultado de Rajchman, 248
- Lindeberg
 - condición de, 260
- Lindeberg, J. W., 220, 260
- Lipschitz, R. O. S., 320
- Lobachevskii, N. I., 339
- Lomnicki, A., 340
- Lyapunov, A. M., 220, 240, 259, 300
- Markov, A. A., 220, 235, 240, 259, 300
- Matrices, 94
 - definidas positivas, 107
 - diagonales, 96
 - escalonadas reducidas, 98
 - invertibles, 97
 - ortogonales, 101
 - polinomio característico, 114
 - simétricas, 100
 - transpuestas, 100
 - triangulares superiores, 97
 - valores propios, 114
 - vectores propios, 114
- Maxwell, J. C., 301
- Medida
 - de probabilidad, 299
 - en espacios de dimensión infinita, 338
 - exterior, 330
 - interior, 330
- Mejor estimador en media cuadrática, 164
- Minkowski, H., 339
- Modelo de Kolmogorov, 339
- Movimiento browniano, 81, 339
- Números normales
 - teorema de Borel sobre los, 317
- Oscilación de una función
 - en un intervalo, 322
 - en un punto, 323
- Paccioli L., 274
- Pascal, B., 219, 271
- Peano, G., 326
- Poincaré, J. H., 301
- Poisson
 - proceso de, 57, 182, 205, 214
 - teorema de, 257
- Poisson, S. D., 220, 235
- Probabilidades numerables, 307
- Problema
 - de la división de apuestas, 271, 274, 278
 - de la medida, 329
 - de la ruina del jugador, 272, 303
 - de los tres jugadores, 302, 315
- Procesos de ramificación, 189
- Radon, J., 332
- Rajchman, 248
- Regla general de la probabilidad total, 183
- Riemann, G. F. B., 320, 321
- Sigma-álgebra de subconjuntos, 298
- Sigma-aditiva
 - función, 298
- Smith, H. J. S., 324
- Steinhaus, H. D., 299, 338
- Stolz, O., 326
- Teoría de la Medida
 - de Borel, 326
 - de Lebesgue, 329
 - surgimiento de la, 319
- Teoría de la Probabilidad moderna
 - surgimiento de la, 297
- Teorema del límite central, 259
- Teoremas límite, 219, 234
- Truncación
 - método de, 240
- Varianza, 68
- Vector aleatorio
 - absolutamente continuo, 15
 - discreto
 - bidimensional, 11
 - n-dimensional, 13
- Volterra, V., 324
- Von Mises, R., 337
- Wiener, N., 37, 299, 338