Probabilidad

con aplicación a retos de la ingeniería moderna

Netzahualcóyotl Castañeda-Leyva
netza.castaneda@edu.uaa.mx

Silvia Rodríguez-Narciso
silvia.rodriguezn@edu.uaa.mx

3 de marzo de 2025

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Espacio de probabilidad

En este capítulo se define formalmente un espacio de probabilidad. Se obtienen las propiedades básicas y se muestran algunos ejemplos.

1.1. Introducción a la probabilidad

El resultado final de un experimento es incierto; no se puede determinar con seguridad lo que ocurrirá. Antes de que se lleve a cabo, el observador del fenómeno de estudio identifica el conjunto de resultados posibles del experimento. Así mismo, entre dichos resultados posibles, se prevén ciertos patrones o tendencias. Aunque el fenómeno de estudio sea aleatorio, habrá eventos más recurrentes o factibles que otros. En este sentido, se construye un modelo matemático de probabilidad del fenómeno aleatorio de interés.

La modelación matemática de los fenómenos con incertidumbre tiene aplicación en todas las ramas del conocimiento. En Medicina por ejemplo, se caracteriza la efectividad de un tratamiento médico, acompañado de la diversidad de los pacientes en términos de su edad, sexo, raza, condición genética, estado de salud o avance de la enfermedad, entre otros estratos. En el contexto industrial, la calidad y durabilidad de los productos manufacturados está basada en modelos de probabilidad. Durante todo el proceso de producción se establecen modelos de probabilidad para identificar las fuentes de variación.

La *probabilidad* o teoría de probabilidad es la rama de las matemáticas que estudia los fenómenos aleatorios. Así mismo, la *Estadística* es la disciplina de la ciencia que colecta, estudia e interpreta los datos de una población. Hay una estrecha relación entre la estadística y la probabilidad. Como parte de las matemáticas, la probabilidad es una ciencia formal, en donde la generación del conocimiento se obtiene por razonamiento del tipo deductivo. En cambio, la estadística es una ciencia fáctica, factual o empírica, en donde el tipo de razonamiento es inductivo. Similar a la biología, física, o sociología, cualquier afirmación estadística puede ser errónea. La Tabla 1.1 describe ejemplos de ciencias fácticas y formales, así como el lugar que guardan la estadística y la probabilidad.

Por ejemplo, para distinguir entre las cualidades deductiva de la probabilidad y deductiva de la estadística, considere una compañía farmacéutica que desarrolla una vacuna contra el Covid-19. El modelo de probabilidad establece una efectividad de la vacuna con una probabilidad fija $0 < p < 1$. En particular, la probabilidad de éxito puede ser $p = 0.9351$. Con los

ciencia fáctica razonamiento inductivo: particular a lo general	**ciencia formal** razonamiento deductivo: general a lo particular
biología	lógica
física	matemáticas
química	álgebra
economía	análisis matemático
historia	geometría
estadística	probabilidad

Tabla 1.1. Ejemplos de ciencias fácticas y formales.

modelo de probabilidad
efectividad de la vacuna: $p = 0.9351$ efectiva para el 93.51 % de los individuos. Porcentaje desconocido y preciso; no es 93.5 % ni 93.52 %.

afirmación estadística
efectividad de 90 %; ¿hay error ó no en la afirmación?
afirmación estadística
efectividad entre 85 % y 90 %; ¿hay error ó no en la afirmación?

Tabla 1.2. Ejemplo de un modelo de probabilidad y de afirmaciones estadísticas; propensas a error.

cual, el 93.51 % de la población vacunada desarrollará anticuerpos mientras que no lo hará el resto; 6.49 %. Este modelo es perfecto, de modo que, si se elige una persona al azar de la población objetivo, la probabilidad de que esta persona desarrolle anticuerpos es 0.9351. Cabe mencionar que en la práctica no se dispone de información de la población en estudio con tal precisión. Por lo cual, sin perder la cualidad perfecta de un modelo matemático, sólo se especifica que la probabilidad del evento de interés es $0 < p < 1$, con p fijo y desconocido. Por otro lado, con los datos obtenidos experimentalmente, la empresa estima que el porcentaje de efectividad es 87.5 %. ¿Hay error ó no en esta afirmación estadística? Del mismo modo, la compañía ofrece otra afirmación estadística expresada en un intervalo de la probabilidad de efectividad de la vacuna entre 85 % y 90 %. ¿Hay error ó no en esta otra afirmación estadística? La Tabla 1.2 describe el modelo de probabilidad de este problema, junto con las dos afirmaciones estadísticas de la empresa farmacéutica sobre la efectividad de la vacuna, las cuales son susceptibles a error.

Con un modelo de probabilidad se obtiene:

1. Una estructura del fenómeno con incertidumbre de interés, que se refleja en conceptos como
   - Espacio muestral
   - Espacio de probabilidad
   - Variable aleatoria y su distribución.

conjunto	totalidad de entes que tienen una propiedad común. Notación usual: A, B, C, Ω
elemento $a \in A$ $b \notin B$	quien está o pertenece al conjunto. Notación usual: a, b, c, ω. a está o pertenece a A b no está o no pertenece a B
vacío $\varnothing$	conjunto que no tiene elementos
cardinalidad $\#A$	tres casos: finita; número de elementos de A, infinita numerable; $\#A = \#\mathbf{N}$ e infinita no numerable; $\#A > \#\mathbf{N}$

Tabla 1.3. Definiciones básicas de conjuntos.

2. El cálculo de posibilidades de los eventos de interés, de modo que identifica tendencias o patrones de ocurrencia. En particular, predice eventos de alta probabilidad mientras que cuando sea baja los descarta.

3. Aplicación en todas las ramas del conocimiento, especialmente en la estadística.

1.2. Álgebra de conjuntos

Esta sección es una introducción al álgebra de conjuntos. En particular se define el concepto de σ-álgebra, la cual es una colección de conjuntos donde se define la medida de probabilidad.

El álgebra de conjunto es útil para definir y calcular la medida de probabilidad en eventos, que son subconjuntos del conjunto de los resultados posibles de un experimento; llamado espacio muestral.

Si bien no hay una definición formal de *conjunto*, este se describe como la totalidad de entes que tienen una propiedad común. Por ejemplo, considere el conjunto A de abejas que viven en cierto panal. La representación $a \in A$, significa que a es un *miembro*, *elemento* o que *pertenece* al conjunto A. En caso contrario, $a \notin A$, significa que el elemento a *no está* o *no pertenece* a dicho conjunto. Usualmente, los conjuntos se denotan con letras mayúsculas; A, B, C, Ω. Mientras que con minúsculas se representan sus elementos; a, b, c, ω.

El conjunto *vacío* es aquel que no tiene elementos. Por ejemplo, vacío es el conjunto de edificios de una altura mayor que 1 000 m. La Tabla 1.3 resume la notación y los conceptos básicos de conjuntos.

Se dice que un conjunto A es de *cardinalidad finita* si hay una asociación *biunívoca* o *biyectiva* entre el conjunto A y el de los números naturales $\{1, \ldots, n\}$; para cierto $n \geq 1$. En este caso, $\#A = n$, denota la *cardinalidad* o número de elementos que contiene A. El conjunto vacío tiene cardinalidad cero: $\#\varnothing = 0$. Por otro lado, el conjunto A es *infinito numerable* si hay una correspondencia biyectiva con el conjunto de los números *naturales* o *enteros positivos*:

$$\mathbf{N} = \{1, 2, 3, \ldots\}.$$

Se denota como $\#A = \#\mathbf{N}$. Ejemplos de conjuntos infinito numerables son: naturales $\mathbf{N}$, enteros $\mathbf{Z}$, enteros pares, enteros impares, enteros no negativos $\mathbf{N} \cup \{0\}$, naturales pares,

finito: $0 \leq \#A < \infty$	
	$0 = \#\varnothing$
	$\boxed{} = \#\{\varnothing\}$
	$\boxed{} = \#\{0\}$
	$\boxed{} = \#\{0, 1, 2, \dots, 9\}$
	$\boxed{} = \#\{-10, -9, \dots, 9, 10\}$

infinito numerable: $\#A = \#\mathbf{N}$	
enteros no negativos	$\mathbf{N} \cup \{0\} = \{0, 1, 2, \dots\}$
enteros	$\mathbf{Z} = \{\dots, -2 - 1, 0, 1, 2, \dots\}$
racionales	$\mathbf{Q} = \left\{\dfrac{m}{n} : m, n \in \mathbf{Z}, n \neq 0\right\}$
	$\mathbf{Z}^2 = \{(m, n) : m, n \in \mathbf{Z}\}$

infinito no numerable: $\#A = \#\mathbf{R} > \#\mathbf{N}$	
reales	$\mathbf{R} = (-\infty, \infty)$
intervalo abierto	$(0, 1) = \{x \in \mathbf{R} : 0 < x < 1\}$
intervalo semiabierto	$(2, 5] = \{x \in \mathbf{R} : 2 < x \leq 5\}$
intervalo cerrado	$[-3, 1] = \{x \in \mathbf{R} : -3 \leq x \leq 1\}$
intervalo semicerrado	$[2, \pi) = \{x \in \mathbf{R} : 2 \leq x < \pi\}$
irracionales	$\mathbf{Q}^c = \{x \in \mathbf{R} : x \notin \mathbf{Q}\}$
plano	$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}$

Tabla 1.4. Ejemplos de conjuntos finitos, infinito numerables o infinito no numerables.

naturales impares, racionales $\mathbf{Q}$, entre otros. Por último, un conjunto A es de *cardinalidad infinita no numerable* si no hay una relación *inyectiva* de A a $\mathbf{N}$. Se denota como $\#A > \#\mathbf{N}$. Por ejemplo, los siguientes conjuntos son infinito no numerables: reales $\mathbf{R}$, irracionales $\mathbf{Q}^c$, cualquier intervalo de longitud positiva o infinita; como $(0, 1)$, $[0, 1]$, $(0, \infty)$ $(\infty, 3)$. La Tabla 1.4 muestra ejemplos de conjuntos finitos, infinito numerables o infinito no numerables, junto con su cardinalidad respectiva. Los espacios en blando pueden ser llenados por el lector.

Sean A y B dos conjuntos. Se dice que A es *subconjunto* de B o B *contiene* a A, si todo elemento de A pertenece a B. En tal caso, se denota con cualquiera de los siguientes maneras:

$$A \subset B, \quad A \subseteq B, \quad B \supset A, \quad B \supseteq A.$$

Los conjuntos A y B son *iguales* $A = B$, siempre que ambos conjuntos compartan sus elementos:

$$A \subset B \quad \text{y} \quad B \subset A.$$

Observe los siguientes ejemplos, con el cuidado de su notación.

$$\begin{array}{ccc}
A \subset A, & \varnothing \subset A, & \varnothing \neq \{\varnothing\}, \\
(0, 1) \subset [0, 1], & [0, 1] \not\subset (0, 1), & (0, 1) \cancel{\in [0, 1]}, \\
\{3\} \subset [3, \pi], & 3 \in [3, \pi], & 3 \cancel{\subset [3, \pi]}.
\end{array}$$

Las tres operaciones básicas de los conjuntos son unión, intersección y complemento. si bien estas operaciones tienen cierta analogía con la suma, producto y resta de los números reales, se aclara que aplican al contexto de conjuntos. La *unión* de los conjuntos A y B es la colección de elementos que están en A o en B:

$$A \cup B \circeq \{x : x \in A \quad \text{o} \quad x \in B\}.$$

La *intersección* de A y B es el conjunto de elementos que pertenecen a A y a B:

$$A \cap B \circeq \{x : x \in A \quad \text{y} \quad x \in B\}.$$

Observe que

$$(A \cap B) \subset A \subset (A \cup B).$$

El álgebra de conjuntos se extiende de forma similar a operaciones finitas e infinitas:

$$\begin{aligned}
\bigcup_{i=1}^{n} A_i &= A_1 \cup \cdots \cup A_n = \{x : x \in A_i \quad \text{para cierto} \quad i = 1, \ldots, n\}, \\
\bigcap_{i=1}^{n} A_i &= A_1 \cap \cdots \cap A_n = \{x : x \in A_i \quad \text{para todo} \quad i = 1, \ldots, n\}, \\
\bigcup_{i=1}^{\infty} A_i &= A_1 \cup A_2 \cup \cdots = \{x : x \in A_i \quad \text{para cierto} \quad i \geq 1\}, \\
\bigcap_{i=1}^{\infty} A_i &= A_1 \cap A_2 \cap \cdots = \{x : x \in A_i \quad \text{para todo} \quad i \geq 1\}.
\end{aligned}$$

Sea A y Ω dos conjuntos tal que $A \subset \Omega$. Respecto de Ω, el conjunto *complemento* de A es

$$A^c \circeq \Omega \setminus A = \{\omega \in \Omega : \omega \notin A\}.$$

El conjunto complemento es relativo o se circunscribe al *espacio muestral* Ω.

Teorema 1.2.1. Las operaciones de conjuntos satisfacen las siguientes propiedades básicas.

1. Propiedad conmutativa

$$\begin{aligned}
A \cup B &= B \cup A, \\
A \cap B &= B \cap A.
\end{aligned}$$

2. Propiedad asociativa

$$\begin{aligned}
A \cup (B \cup C) &= (A \cup B) \cup C = (A \cup C) \cup B = A \cup B \cup C, \\
A \cap (B \cap C) &= (A \cap B) \cap C = (A \cap C) \cap B = A \cap B \cap C.
\end{aligned}$$

3. Propiedad distributiva

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \tag{1.1}$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \tag{1.2}$$

$$A \cup \left(\bigcap_{i=1}^{n} B_i \right) = \bigcap_{i=1}^{n} (A \cup B_i), \qquad A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} (A \cap B_i).$$
$$A \cup \left(\bigcap_{i=1}^{\infty} B_i \right) = \bigcap_{i=1}^{\infty} (A \cup B_i), \qquad A \cap \left(\bigcup_{i=1}^{\infty} B_i \right) = \bigcup_{i=1}^{\infty} (A \cap B_i).$$

4.
$$(A^c)^c = A.$$

Demostración. Se verificarán algunas de las propiedades de este teorema: la primera conmutativa junto con la segunda distributiva. Se deja al lector la comprobación del resto de las propiedades; véanse los ejercicios 1.2-1.5.
1. Para probar
$$A \cup B = B \cup A,$$
se verifica que ambos conjuntos comparten los mismos elementos:
$$A \cup B \subset B \cup A \quad \text{y} \quad B \cup A \subset A \cup B.$$

Paso ($\subset$). Si $x \in A \cup B$, entonces

$$\begin{aligned} &x \in A \quad \text{o} \quad x \in B, \\ &x \in B \quad \text{o} \quad x \in A, \\ &x \in B \cup A. \end{aligned}$$

Así que todo elemento de $A \cup B$ está en $B \cup A$.
Paso ($\supset$). Se procede en sentido contrario.
3. Ahora se verificará la propiedad distributiva (1.2).
Paso ($\subset$). Si $\omega \in A \cap (B \cup C)$, entonces

$$\begin{aligned} &\omega \in A \cap (B \cup C), \\ &\omega \in A \quad \text{y} \quad \omega \in (B \cup C), \\ &\omega \in A \quad \text{y} \quad \omega \in B \quad \text{u} \quad \omega \in C, \\ &\omega \in A \quad \text{y} \quad \omega \in B \quad \text{u} \quad \omega \in A \quad \text{y} \quad \omega \in C, \\ &\omega \in A \cap B \quad \text{u} \quad \omega \in A \cap C, \\ &\omega \in (A \cap B) \cup (A \cap C). \end{aligned}$$

Paso ($\supset$). Se verifica al proceder en sentido contrario. $\square$

Ejemplo 1.2.2. Verifique las dos propiedades distributivas de conjuntos (1.1)-(1.2), cuando

$$A = (0,3), \quad B = [2,5), \quad C = (1,4].$$

La Figura 1.1 muestra la gráfica de estos tres intervalos. Note que

$$B \cap C = [2,5) \cap (1,4] = [2,4].$$

Aquí se usó

$$\begin{array}{lll} 2 \in [2,5) \quad \text{y} \quad 2 \in [1,4]; & \text{que implica} & 2 \in [2,4], \\ 4 \in [2,5) \quad \text{y} \quad 4 \in (1,4]; & \text{que implica} & 4 \in [2,4]. \end{array}$$

Así

$$A \cup (B \cap C) = (0,3) \cup [2,4] = (0,4].$$

Por otro lado

$$\begin{aligned} A \cup B &= (0,3) \cup [2,5) = (0,5), \\ A \cup C &= (0,3) \cup (1,4] = (0,4]. \end{aligned}$$

Por lo tanto, se verifica (1.1).

$$(A \cup B) \cap (A \cup C) = (0,5) \cap (0,4] = (0,4] = A \cup (B \cap C).$$

Así mismo, para verificar (1.2), considere

$$B \cup C = [2,5) \cup (1,4] = (1,5).$$

Así

$$A \cap (B \cup C) = (0,3) \cap (1,5) = (1,3).$$

Aquí se usó

$$\begin{array}{lll} 1 \in (0,3) \quad \text{y} \quad 1 \notin (1,5); & \text{que implica} & 1 \notin (1,3), \\ 3 \notin (0,3) \quad \text{y} \quad 3 \in (1,5); & \text{que implica} & 3 \notin (1,3). \end{array}$$

Por otro lado

$$\begin{aligned} A \cap B &= (0,3) \cap [2,5) = [2,3), \\ A \cap C &= (0,3) \cap (1,4] = (1,3). \end{aligned}$$

Aquí se usó

$$\begin{array}{lll} 2 \in (0,3) \quad \text{y} \quad 2 \in [2,5); & \text{que implica} & 2 \in [2,3), \\ 3 \notin (0,3) \quad \text{y} \quad 3 \in [2,5); & \text{que implica} & 3 \notin [2,3), \\ 1 \in (0,3) \quad \text{y} \quad 1 \notin (1,4]; & \text{que implica} & 1 \notin (1,3), \\ 3 \notin (0,3) \quad \text{y} \quad 3 \in (1,4]; & \text{que implica} & 3 \notin (1,3). \end{array}$$

Por lo tanto

$$(A \cap B) \cup (A \cap C) = [2,3) \cup (1,3) = (1,3) = A \cap (B \cup C).$$

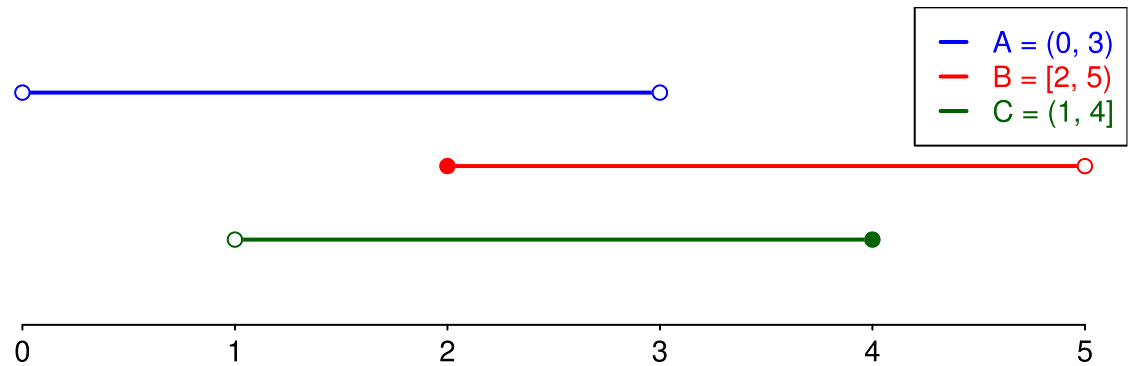


Figura 1.1. Gráfica de tres intervalos; Ejemplo 1.2.2.

Una sucesión de conjuntos $\{A_n\}_{n\geq 1}$, es *creciente* si

$$A_1 \subset A_2 \subset \cdots \subset A = \bigcup_{n=1}^{\infty} A_n.$$

Se denota como $A_n \uparrow A$; cuando $n \to \infty$, donde A es la unión de los conjuntos. Del mismo modo, se escribe $A_n \downarrow A$; cuando $n \to \infty$, si $\{A_n\}_{n\geq 1}$ es una sucesión *decreciente* de conjuntos, la cual decrece a la intersección:

$$A_1 \supset A_2 \supset \cdots = A = \bigcap_{n=1}^{\infty} A_n.$$

Ejemplo 1.2.3. Se describen ejemplos de sucesiones monótonas de conjuntos.

1. La sucesión de intervalos $A_n = (-\infty, n)$; para $n \geq 1$, crece a la recta real: $A_n \uparrow \mathbf{R}$; cuando $n \to \infty$. De hecho

$$\begin{aligned}
A_1 &= (-\infty, 1) \subset (-\infty, 2) = A_2,\\
A_2 &= (-\infty, 2) \subset (-\infty, 3) = A_3,\\
A_3 &= (-\infty, 3) \subset (-\infty, 4) = A_4,\\
&\vdots\\
A_n &= (-\infty, n) \subset (-\infty, n+1) = A_{n+1} \subset \cdots \subset \mathbf{R}.
\end{aligned}$$

   Para verificar $\bigcup_{n=1}^{\infty} A_n = \mathbf{R}$, aplique la propiedad arquimediana de los números reales. Si $x \in \mathbf{R}$, entonces existe un entero $n > 0$ tal que $x < n$. De hecho, considere $n = \text{máx}\{[x], 0\} + 1$, donde $[\,\cdot\,]$ denota la función *parte entera*.

2. La sucesión $A_n = (-\infty, -n)$; $n \geq 1$, decrece al conjunto vacío: $A_n \downarrow \varnothing$; cuando $n \to \infty$. Los conjuntos $A_1, A_2, \ldots,$ no tienen elementos en común. En el idioma Español se dice también “no tienen ningún elemento en común” o “no tienen elemento en común alguno”.

Note que

$$\begin{aligned}
A_2 &= (-\infty,-2) \subset (-\infty,-1) = A_1,\\
A_3 &= (-\infty,-3) \subset (-\infty,-2) = A_2,\\
A_4 &= (-\infty,-4) \subset (-\infty,-3) = A_3,\\
&\vdots\\
\varnothing \subset \cdots \subset A_{n+1} &= (-\infty,-n-1) \subset (-\infty,-n) = A_n.
\end{aligned}$$

Sea $x \in \mathbf{R}$. Si $x \geq -1$, entonces $x \notin A_1$. Ahora bien, si $x < -1$, existe $n \geq 2$ tal $x \geq -n$; como $n = [-x-1]$. Así que $x \notin A_n$. En cualquier caso, se tiene $x \notin \bigcap_{n=1}^{\infty} A_n$. Por lo tanto, esta intersección es vacía.

3. Si $A_n = [0, 1 + 1/n)$; para $n \geq 1$, entonces $A_n \downarrow [0,1]$; cuando $n \to \infty$. De hecho

$$\begin{aligned}
A_2 &= \left[0, 1+\frac{1}{2}\right) = \left[0, \frac{3}{2}\right) \subset [0,2) = \left[0, 1+\frac{1}{1}\right) = A_1,\\
A_3 &= \left[0, 1+\frac{1}{3}\right) = \left[0, \frac{4}{3}\right) \subset \left[0, \frac{3}{2}\right) = A_2,\\
A_4 &= \left[0, 1+\frac{1}{4}\right) = \left[0, \frac{5}{4}\right) \subset \left[0, \frac{4}{3}\right) = A_3,\\
&\vdots
\end{aligned}$$

Esta sucesión de conjuntos decrece a $A = \bigcap_{n=1}^{\infty} A_n = [0,1]$. El número extremo 1 es elemento de la intersección A, puesto que pertenece a cada conjunto de la sucesión:

$$\begin{aligned}
&1 \in A_n; \quad \text{para todo} \quad n \geq 1,\\
&1 \in \left[0, 1+\frac{1}{n}\right),\\
&1 < 1+\frac{1}{n},\\
&0 < \frac{1}{n}.
\end{aligned}$$

4. Si $A_n = [0, 1 - 1/n)$; para $n \geq 1$, entonces $A_n \uparrow [0,1)$; cuando $n \to \infty$. De hecho

$$\begin{aligned}
A_2 &= \left[0, 1-\frac{1}{2}\right) = \left[0, \frac{1}{2}\right) \supset \varnothing = [0,0) = \left[0, 1-\frac{1}{1}\right) = A_1,\\
A_3 &= \left[0, 1-\frac{1}{3}\right) = \left[0, \frac{2}{3}\right) \supset \left[0, \frac{1}{2}\right) = A_2,\\
A_4 &= \left[0, 1-\frac{1}{4}\right) = \left[0, \frac{3}{4}\right) \supset \left[0, \frac{2}{3}\right) = A_3,\\
&\vdots
\end{aligned}$$

El número 1 no pertenece a la unión A, pues no está en ningún conjunto de la sucesión:

$$
\begin{aligned}
1 &\notin A_n; \quad \text{para todo} \quad n \geq 1,\\
1 &\in \left[0, 1 - \frac{1}{n}\right)^c,\\
1 &\in (-\infty, 0) \cup \left[1 - \frac{1}{n}, \infty\right),\\
1 &\geq 1 - \frac{1}{n},\\
\frac{1}{n} &\geq 0.
\end{aligned}
$$

Los conjuntos A y B son *excluyentes* o *ajenos*, si

$$A \cap B = \varnothing.$$

Una sucesión $\{B_n\}_{n\geq 1}$ de eventos de Ω es *excluyente* si son excluyentes por pares:

$$B_i \cap B_j = \varnothing; \quad i \neq j, \quad i, j \geq 1.$$

Esta sucesión es además una *partición* de Ω si su unión es el propio espacio muestral:

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

En tal caso, a los conjuntos excluyentes $B_1, B_2, \ldots,$ se les dice *estratos* de Ω. Por ejemplo, para cualquier conjunto $B \subset \Omega$, los conjuntos B y B^c forman una partición finita de Ω; puesto que son eventos excluyentes y su unión es Ω:

$$B \cap B^c = \varnothing \quad \text{y} \quad B \cup B^c = \Omega.$$

Así mismo, se dice que una sucesión de conjuntos $\{A_n\}_{n\geq 1}$ es una *partición* de un conjunto A, si es una colección de conjuntos excluyentes y su unión es A:

$$A_i \cap A_j = \varnothing; \quad i \neq j, \quad i, j \geq 1 \quad \text{y} \quad \bigcup_{n=1}^{\infty} A_n = A.$$

Por ejemplo, B y B^c inducen una partición de cualquier otro conjunto A:

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c),$$

con

$$(A \cap B) \cap (A \cap B^c) = (A \cap A) \cap (B \cap B^c) = A \cap \varnothing = \varnothing.$$

1.2.1. Leyes de De Morgan

Las leyes de De Morgan se refieren a la fórmula del complemento de una unión o intersección de conjuntos, ya sea finita, infinita numerable o, incluso, no numerable. En esta sección se describen sus versiones finito e infinito numerable.

Dado dos subconjuntos A y B de Ω, se tiene

$$(A \cup B)^c = A^c \cap B^c, \tag{1.3}$$
$$(A \cap B)^c = A^c \cup B^c. \tag{1.4}$$

Así mismo

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c, \qquad \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c, \tag{1.5}$$
$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c, \qquad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c. \tag{1.6}$$

Para verificar la igualdad (1.4), debe comprobarse

$$(A \cap B)^c \subset A^c \cup B^c \quad \text{y} \quad A^c \cup B^c \subset (A \cap B)^c.$$

Paso $\subset$. Tome un elemento arbitrario $\omega \in (A \cap B)^c$, para luego comprobar su pertenencia al conjunto $A^c \cup B^c$:

$$\begin{aligned} \omega \in (A \cap B)^c \quad &\Rightarrow \quad \omega \notin A \cap B, \\ &\Rightarrow \quad \omega \notin A \quad \text{u} \quad \omega \notin B, \\ &\Rightarrow \quad \omega \in A^c \quad \text{u} \quad \omega \in B^c, \\ &\Rightarrow \quad \omega \in A^c \cup B^c. \end{aligned}$$

Por lo tanto

$$(A \cap B)^c \subset A^c \cup B^c.$$

Paso $\supset$. Considere $\omega \in A^c \cup B^c$. Al proceder en sentido contrario del paso anterior, se deduce que $\omega \in (A \cap B)^c$. Por lo tanto

$$A^c \cup B^c \subset (A \cap B)^c$$

y

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

De modo similar, se comprueba la relación (1.3); véase el Ejercicio 1.7. Otra idea es aplicar la segunda ley de De Morgan (1.4):

$$\begin{aligned} A \cup B = (A^c)^c \cup (B^c)^c = (A^c \cap B^c)^c, \\ (A \cup B)^c = [(A^c \cap B^c)^c]^c = A^c \cap B^c. \end{aligned}$$

La versión finita de las leyes de De Morgan (1.5) se verifica de manera similar.

$$\begin{aligned}\left(\bigcap_{i=1}^{n} A_i\right)^c &= \left\{\omega \in \Omega : \omega \notin \bigcap_{i=1}^{n} A_i\right\} \\ &= \{\omega \in \Omega : \omega \notin A_i, \quad \text{para cierto} \quad i = 1, \ldots, n\} \\ &= \{\omega \in \Omega : \omega \in A_i^c, \quad \text{para cierto} \quad i = 1, \ldots, n\} \\ &= \bigcup_{i=1}^{n} A_i^c.\end{aligned}$$

Así mismo

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} (A_i^c)^c = \left(\bigcap_{i=1}^{n} A_i^c\right)^c,$$

$$\left[\bigcup_{i=1}^{n} A_i\right]^c = \left[\left(\bigcap_{i=1}^{n} A_i^c\right)^c\right]^c = \bigcap_{i=1}^{n} A_i^c.$$

Se deja al lector comprobar la versión infinita numerable de las leyes de Morgan (1.6); véase Ejercicio 1.8.

Por ejemplo, si $B = (-\infty, 7)$ y $C = (2, \infty)$, se valida (1.3)-(1.4):

$$\begin{aligned}(B \cup C)^c &= ((-\infty, 7) \cup (2, \infty))^c = (-\infty, \infty)^c = \varnothing = (-\infty, 2] \cap [7, \infty) = B^c \cap C^c, \\ (B \cap C)^c &= ((-\infty, 7) \cap (2, \infty))^c = (2, 7)^c = (-\infty, 2] \cup [7, \infty) = B^c \cup C^c.\end{aligned}$$

La Figura 1.2 muestra los intervalos involucrados en este ejemplo.

Ahora considere el siguiente ejemplo de intervalos. Sea el conjunto

$$A = \{x \in \mathbf{R} : |x - 5| < 4\}.$$

Este conjunto se denota informalmente como $|x - 5| < 4$. Su solución es

$$1 < x < 9.$$

Defina los siguientes conjuntos

$$B = \{x \in \mathbf{R} : x > 1\} = (1, \infty), \quad C = \{x \in \mathbf{R} : x < 9\} = (-\infty, 9).$$

Así que

$$A = (1, 9) = (1, \infty) \cap (-\infty, 9) = B \cap C.$$

Por leyes de De Morgan, se tiene

$$A^c = (B \cap C)^c = B^c \cup C^c = (1, \infty)^c \cup (-\infty, 9)^c = (-\infty, 1] \cup [9, \infty) = A^c$$

y

$$\begin{aligned}(B \cup C)^c &= ((1, \infty) \cup (-\infty, 9))^c = (-\infty, \infty)^c = \varnothing \\ &= (-\infty, 1] \cap [9, \infty) = (1, \infty)^c \cap (-\infty, 9)^c \\ &= B^c \cap C^c.\end{aligned}$$

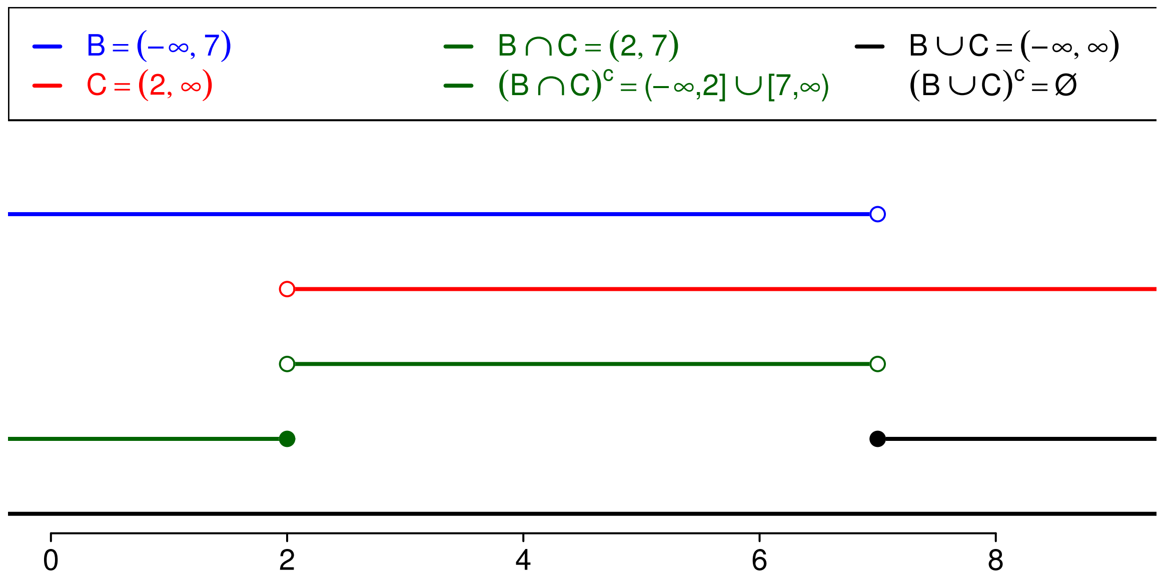


Figura 1.2. Gráfica de intervalos que describe las leyes de De Morgan.

Ahora asuma

$$A = \left(0, \frac{1}{8}\right), \quad B = \left(\frac{1}{8}, \frac{1}{4}\right), \quad C = \left(\frac{1}{2}, \frac{3}{4}\right).$$

Esta es una colección de tres eventos excluyentes. Así

$$A \cap B \cap C = \varnothing$$

y

$$(A \cap B \cap C)^c = \varnothing^c = \Omega = \mathbf{R}.$$

Por otro lado

$$A \cup B \cup C = \left(0, \frac{1}{8}\right) \cup \left(\frac{1}{8}, \frac{1}{4}\right) \cup \left(\frac{1}{2}, \frac{3}{4}\right),$$

y

$$\begin{aligned}(A \cup B \cup C)^c &= A^c \cap B^c \cap C^c \\ &= \left(0, \frac{1}{8}\right)^c \cap \left(\frac{1}{8}, \frac{1}{4}\right)^c \cap \left(\frac{1}{2}, \frac{3}{4}\right)^c \\ &= (-\infty, 0] \cup \left\{\frac{1}{8}\right\} \cup \left[\frac{1}{4}, \frac{1}{2}\right] \cup \left[\frac{3}{4}, \infty\right).\end{aligned}$$

Para la última igualdad apóyese en una gráfica.

Si el espacio muestral cambia a $\Omega = (0, 1)$, entonces el complemento de los conjuntos anteriores son relativos al nuevo espacio muestral. Por ejemplo

$$\begin{aligned} A^c &= \left(0, \frac{1}{8}\right)^c = \left[\frac{1}{8}, 1\right), \\ B^c &= \left(\frac{1}{8}, \frac{1}{4}\right)^c = \left(0, \frac{1}{8}\right] \cup \left[\frac{1}{4}, 1\right), \\ C^c &= \left(\frac{1}{2}, \frac{3}{4}\right)^c = \left(0, \frac{1}{2}\right] \cup \left[\frac{3}{4}, 1\right). \end{aligned}$$

Así mismo

$$\begin{aligned} (A \cap B \cap C)^c &= \varnothing^c = \Omega = (0,1), \\ A \cup B \cup C &= \left(0, \frac{1}{8}\right) \cup \left(\frac{1}{8}, \frac{1}{4}\right) \cup \left(\frac{1}{2}, \frac{3}{4}\right) \end{aligned}$$

y

$$\begin{aligned} (A \cup B \cup C)^c &= A^c \cap B^c \cap C^c \\ &= \left(0, \frac{1}{8}\right)^c \cap \left(\frac{1}{8}, \frac{1}{4}\right)^c \cap \left(\frac{1}{2}, \frac{3}{4}\right)^c \\ &= \left[\frac{1}{8}, 1\right) \cap \left(\left(0, \frac{1}{8}\right] \cup \left[\frac{1}{4}, 1\right)\right) \cap \left(\left(0, \frac{1}{2}\right] \cup \left[\frac{3}{4}, 1\right)\right) \\ &= \left\{\frac{1}{8}\right\} \cup \left[\frac{1}{4}, \frac{1}{2}\right] \cup \left[\frac{3}{4}, 1\right). \end{aligned}$$

El siguiente ejemplo, además de ilustrar el álgebra de conjuntos, es una introducción al cálculo de probabilidad de eventos.

Ejemplo 1.2.4. Considere el experimento de lanzar una moneda 3 veces y registre su cara superior al caer. El conjunto de resultados posibles del experimento es

$$\begin{aligned} \Omega &= \{aaa, aas, asa, ass, saa, sas, ssa, sss\} \\ &= \{(x, y, z) : x, y, z = a \text{ ó } s\}. \end{aligned}$$

La Figura 1.3 muestra la gráfica del espacio muestral Ω. Cada vértice del cubo es un evento simple. La Tabla 1.16, más adelante, muestra también el espacio muestral en un diagrama de árbol, donde cada rama es un evento simple. Aquí las letras E y F sustituyen respectivamente a a y s. El total de eventos simples o resultados posibles del experimento es $\#\Omega = 8$. Con lo que Ω tiene $2^8 = 256$ subconjuntos o eventos; véase Ejercicio 1.8.11. Cada uno de los eventos tiene una probabilidad asignada. Por ejemplo, defina los eventos

$$\begin{aligned} &A : \text{hay al menos dos soles} = \{ass, sas, ssa, sss\}, \\ &B : \text{hay a lo más dos soles} = \{aaa, aas, asa, ass, saa, sas, ssa\}, \\ &C : \text{hay exactamente dos águilas} = \{aas, asa, saa\}, \\ &D : \text{son águila los primeros dos lanzamientos} = \{aaa, aas\}, \\ &E : \text{águila en el primer ensayo} = \{aaa, aas, asa, ass\}. \end{aligned}$$

Los eventos A y C son excluyentes, puesto que no tienen elementos en común:

$$A \cap C = \{ass, sas, ssa, sss\} \cap \{aas, asa, saa\} \ = \varnothing.$$

Así mismo, son excluyentes los eventos A y D:

$$A \cap D = \{ass, sas, ssa, sss\} \cap \{aaa, aas\} \ = \varnothing.$$

En cambio, los eventos A y B no son excluyentes:

$$A \cap B : \text{hay dos soles} = \{ass, sas, ssa\} \neq \varnothing.$$

Tampoco son excluyentes C y D:

$$C \cap D = \{aas, asa, saa\} \cap \{aaa, aas\} = \{aas\} \neq \varnothing.$$

Asuma que la probabilidad un evento de Ω es proporcional al número de eventos simples que contiene:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{8}; \quad \text{para} \quad A \subset \Omega.$$

En particular, cada evento simple tiene la misma probabilidad de ocurrir:

$$\begin{aligned} P(aaa) &= P(aas) = P(asa) = P(asa) \\ &= P(saa) = P(sas) = P(ssa) = P(sss) \\ &= \frac{1}{8} = 0.125 = 12.5\,\%. \end{aligned}$$

Note que

$$\begin{aligned} P(C \cup D) &= P(C) + P(D) - P(C \cap D) \\ &= \frac{3}{8} + \frac{2}{8} - \frac{1}{8} = \frac{4}{8} \\ &= 0.5 = 50\,\%. \end{aligned}$$

Se confirma la probabilidad de este evento al observar que $C \cup D$ tiene 4 eventos simples:

$$C \cup D = \{aas, asa, saa\} \cup \{aaa, aas\} = \{aaa, aas, asa, saa\}.$$

Cuando dos eventos son excluyentes, es más fácil calcular la probabilidad de la unión:

$$\begin{aligned} P(A \cup C) &= P(A) + P(C) - P(A \cap C) \\ &= P(A) + P(C) - P(\varnothing) \\ &= P(A) + P(C) \\ &= \frac{4}{8} + \frac{3}{8} = \frac{7}{8} \\ &= 0.875 = 87.5\,\%. \end{aligned}$$

$$
\begin{tikzcd}
 & ssa \arrow[rr] \arrow[dl] & & saa \arrow[dl] \arrow[dd] \\
asa \arrow[rr] & & aaa & \\
 & sss \arrow[uu] \arrow[rr] \arrow[dl] & & sas \arrow[dl] \\
ass \arrow[uu] \arrow[rr] & & aas \arrow[uu] &
\end{tikzcd}
$$

Figura 1.3. Gráfica del espacio muestral Ω, en el lanzamiento de una moneda tres veces. Cada vértice es un evento simple; Ejemplo 1.2.4.

Se confirma la probabilidad de este evento al observar que $A \cup C$ tiene 7 eventos simples:

$$
\begin{aligned}
A \cup C &= \{ass, sas, ssa, sss\} \cup \{aas, asa, saa\} \\
&= \{aas, asa, ass, saa, sas, ssa, sss\}.
\end{aligned}
$$

Note que

$$
\begin{aligned}
E^c &: \text{no es águila en el primer ensayo} \\
&: \text{sol en el primer ensayo} \\
&= \{saa, sas, ssa, sss\}.
\end{aligned}
$$

Los eventos E y E^c forman una partición del espacio muestral Ω, puesto que son excluyentes y su unión es Ω. Así mismo, los eventos $A \cap E$ y $A \cap E^c$ son una partición del evento A:

$$
\begin{aligned}
A \cap E &= \{ass, sas, ssa, sss\} \cap \{aaa, aas, asa, ass\} = \{ass\}, \\
A \cap E^c &= \{ass, sas, ssa, sss\} \cap \{saa, sas, ssa, sss\} = \{sas, ssa, sss\}, \\
(A \cap E) \cup (A \cap E^c) &= \{ass\} \cup \{sas, ssa, sss\} = \{ass, sas, ssa, sss\} = A, \\
(A \cap E) \cap (A \cap E^c) &= \{ass\} \cap \{sas, ssa, sss\} = \varnothing.
\end{aligned}
$$

Ejemplo 1.2.5. Sea el experimento de lanzar un dado dos veces y registrar el número de su cara superior. El conjunto de resultados posibles del experimento es

$$\Omega = \{(x, y) : x, y = 1, \ldots, 6\}.$$

La Tabla 1.5 representa al espacio muestral Ω como una matriz de 6×6, en donde el reglón indica el resultado del primer ensayo mientras que del segundo ensayo lo marca la columna. Note que $\#\Omega = 36$. Así mismo, el espacio muestral Ω tiene $2^{36} > 6.87 \times 10^{10}$ subconjuntos. Cada uno de estos subconjuntos tiene asignada una probabilidad, según la fórmula

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{36} \quad \text{para} \quad A \subset \Omega.$$

Se calculará la probabilidad de algunos eventos de Ω. Los eventos simples son equiprobables:

$$\begin{aligned} P((1,1)) &= P((1,2)) = P((1,3)) = \cdots = P((6,5)) = P((6,6)) \\ &= \frac{1}{\#\Omega} = \frac{1}{36} = 0.0277 = 2.77\,\%. \end{aligned}$$

Defina los eventos

$$\begin{aligned} A &: \text{resultado del primer lanzamiento es } 3, \\ B &: \text{resultado del primer lanzamiento es } 5. \end{aligned}$$

Los eventos A y B son excluyentes, pues no tienen elementos en común: $A \cap B = \varnothing$; véase la Tabla 1.5. Ambos eventos tienen la misma probabilidad

$$\begin{aligned} P(A) &= \frac{\#A}{\#\Omega} = \frac{6}{36} = \frac{1}{6} = 0.1667 = 16.67\,\%, \\ P(B) &= \frac{\#B}{\#\Omega} = \frac{6}{36} = \frac{1}{6} = 0.1667 = 16.67\,\%. \end{aligned}$$

Ahora considere los eventos

$$\begin{aligned} C &: \text{suma es mayor que } 7 = \{(2,6),(3,5),(3,6),(4,4),(4,5),(4,6),(5,3),(5,4), \\ &\qquad\qquad (5,5),(5,6),(6,2),(6,3),(6,4),(6,5),(6,6)\}, \\ C_8 &: \text{suma es } 8 = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}, \\ C_9 &: \text{suma es } 9 = \{(3,6),(4,5),(5,4),(6,3)\}, \\ C_{10} &: \text{suma es } 10 = \{(4,6),(5,5),(6,4)\}, \\ C_{11} &: \text{suma es } 11 = \{(5,6),(6,5)\}, \\ C_{12} &: \text{suma es } 12 = \{(6,6)\}. \end{aligned}$$

La colección de eventos $C_8, C_9, C_{10}, C_{11}, C_{12}$ es una partición de C, pues es excluyente y su unión resulta C:

$$\begin{aligned} C_8 \cap C_9 &= C_8 \cap C_{10} = C_8 \cap C_{11} = C_8 \cap C_{12} = C_9 \cap C_{10} = C_9 \cap C_{11} \\ &= C_9 \cap C_{12} = C_{10} \cap C_{11} = C_{10} \cap C_{12} = C_{11} \cap C_{12} = \varnothing \end{aligned}$$

y

$$C = C_8 \cup C_9 \cup C_{10} \cup C_{11} \cup C_{12}.$$

Por lo tanto

$$\begin{aligned} P(C) &= P(C_8 \cup C_9 \cup C_{10} \cup C_{11} \cup C_{12}) \\ &= P(C_8) + P(C_9) + P(C_{10}) + P(C_{11}) + P(C_{12}) \\ &= \frac{\#C_8}{\#\Omega} + \frac{\#C_9}{\#\Omega} + \frac{\#C_{10}}{\#\Omega} + \frac{\#C_{11}}{\#\Omega} + \frac{\#C_{12}}{\#\Omega} \\ &= \frac{5+4+3+2+1}{36} = \frac{15}{36}. \end{aligned}$$

Con los eventos A y C, verifique las leyes de De Morgan (1.3)-(1.4). Apóyese en la Tabla 1.5.

$$\begin{aligned}
(A\cup C)^c = \quad & \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
& (2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
& (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
& (4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
& (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
& (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}^c\\
= \quad & \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
& (2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
& (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
& (4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
& (5,1)(5,2),(5,3),(5,4),(5,5),(5,6),\\
& (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\\
= \quad & A^c\cap C^c
\end{aligned}$$

y

$$\begin{aligned}
(A\cap C)^c = \quad & \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\
& (2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\
& (3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\
& (4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\
& (5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\
& (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\\
= \quad & A^c\cup C^c.
\end{aligned}$$

Ω	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	$\cancel{(2,6)}$
3	(3,1)	(3,2)	(3,3)	(3,4)	$\cancel{(3,5)}$	$\cancel{(3,6)}$
4	(4,1)	(4,2)	(4,3)	$\cancel{(4,4)}$	$\cancel{(4,5)}$	$\cancel{(4,6)}$
5	(5,1)	(5,2)	$\cancel{(5,3)}$	$\cancel{(5,4)}$	$\cancel{(5,5)}$	$\cancel{(5,6)}$
6	(6,1)	$\cancel{(6,2)}$	$\cancel{(6,3)}$	$\cancel{(6,4)}$	$\cancel{(6,5)}$	$\cancel{(6,6)}$

$\cancel{(1,1)}$	$\cancel{(1,2)}$	$\cancel{(1,3)}$	$\cancel{(1,4)}$	$\cancel{(1,5)}$	$\cancel{(1,6)}$
$\cancel{(2,1)}$	$\cancel{(2,2)}$	$\cancel{(2,3)}$	$\cancel{(2,4)}$	$\cancel{(2,5)}$	(2,6)
$\cancel{(3,1)}$	$\cancel{(3,2)}$	$\cancel{(3,3)}$	$\cancel{(3,4)}$	(3,5)	(3,6)
$\cancel{(4,1)}$	$\cancel{(4,2)}$	$\cancel{(4,3)}$	(4,4)	(4,5)	(4,6)
$\cancel{(5,1)}$	$\cancel{(5,2)}$	(5,3)	(5,4)	(5,5)	(5,6)
$\cancel{(6,1)}$	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

A B C A^c C^c

Tabla 1.5. Espacio muestral Ω del lanzamiento de un dado dos veces, así como algunos eventos: A amarillo, B rosa, C raya roja, A^c verde y C^c raya azul; Ejemplo 1.2.5.

1.2.2. σ-álgebra

En esta sección se describe el concepto de σ-álgebra, así como sus propiedades básicas.

La medida probabilidad se define en una familia $\mathcal{A}$ de subconjuntos del espacio muestral Ω, llamada σ-álgebra. Es deseable que esta familia de conjuntos sea lo suficientemente rica en eventos.

Definición 1.2.6. Sea $\mathcal{A}$ una familia de subconjuntos de un conjunto no vacío Ω. Se dice que $\mathcal{A}$ es una *σ-álgebra* o un *σ-campo*, si satisface las siguientes propiedades:

(i) $\Omega \in \mathcal{A}$

(ii) $A^c \in \mathcal{A}$, siempre que $A \in \mathcal{A}$

(iii) $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, siempre que $A_n \in \mathcal{A}$, para todo $n = 1, 2, \ldots$

E *espacio muestral* Ω es el conjunto de resultados posibles de un experimento. Los elementos de $\mathcal{A}$ son subconjuntos de Ω, llamados *eventos*. Si $\omega \in \Omega$, entonces $\{\omega\}$ es un *evento simple* siempre que $\{\omega\} \in \mathcal{A}$. Si bien no es necesario, regularmente las σ-álgebras contienen a todos los elementos de Ω, de modo que sean eventos simples.

Las propiedades (i) e (iii) se pueden reemplazar por sus equivalentes respectivos:

(i') $\varnothing \in \mathcal{A}$

(iii') $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$, siempre que $A_n \in \mathcal{A}$, para todo $n = 1, 2, \ldots$

En efecto, al considerar (ii), las afirmaciones (i) e (i') son equivalentes:

$$\varnothing = \Omega^c \in \mathcal{A} \quad \Leftrightarrow \quad \Omega = \varnothing^c \in \mathcal{A}.$$

Así mismo, por (ii) y las leyes de De Morgan, las afirmaciones (iii) e (iii') son equivalentes: si $A_n \in \mathcal{A}$; para $n = 1, 2, \ldots,$ entonces $A_n^c \in \mathcal{A}$; para todo n. Por lo que

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{A},$$
$$\bigcap_{n=1}^{\infty} A_n = \left[\left(\bigcap_{n=1}^{\infty} A_n\right)^c\right]^c \in \mathcal{A}.$$

El recíproco se obtiene al operar en sentido contrario.

Un espacio muestral Ω puede tener tiene varias σ-álgebras. La riqueza de una σ-álgebra se determina por la diversidad de eventos del espacio muestral. Lo que refleja la perspectiva de la información sobre los resultados del experimento. Toda σ-álgebra incluye tanto al conjunto vacío $\varnothing$ como al espacio muestral Ω.

Cualquier colección de subconjuntos de Ω induce una σ-álgebra. Si $\mathcal{A}$ es una colección de subconjuntos de Ω, no necesariamente una σ-álgebra, su σ-álgebra *generada* es la mínima σ-álgebra en Ω que la contiene:

$$\sigma(\mathcal{A}) = \text{mínima } \sigma\text{-álgebra que contiene a } \mathcal{A}.$$

Si en particular $\mathcal{A}$ es una σ-álgebra, entonces $\sigma(\mathcal{A}) = \mathcal{A}$.

Ejemplo 1.2.7. En este ejemplo se describen algunas σ-álgebras clásicas.

1. La σ-álgebra más pequeña de un espacio muestral Ω es
$$\mathcal{A} = \{\varnothing, \Omega\}.$$
Esta es la σ-álgebra menos informativa posible, puesto que sólo especifica que se llevó a cabo el experimento. Imagine por ejemplo que se lanza una moneda y que se tapa su cara superior al caer, de modo que no se observa el resultado del experimento. Entonces, el único evento no vacío de este experimento es el propio espacio muestral: $\Omega = \{$águila, sol$\}$. Note además que $\mathcal{A}$ no contiene a los conjuntos $\{$águila$\}$ y $\{$sol$\}$. Por lo que formalmente no son eventos simples.

2. La σ-álgebra generada por un conjunto no vacío $A \subset \Omega$, es
$$\mathcal{A} = \sigma(A) = \{\varnothing, A, A^c, \Omega\}.$$
Aún pobre, esta es una σ-álgebra un poco más informativa que la anterior. Salvo el caso $\#\Omega = 2$, la σ-álgebra $\mathcal{A}$ no contiene a sus eventos simples. Con ella se especifican los eventos de interés en términos del evento A: ocurrió o no dicho evento. Por ejemplo, en una temporada de lluvias, considere el evento A de que se desborda una presa. Entonces, A^c significa que no se desborda. Con una pobre gama de eventos, no se puede matizar con otros tipo de eventos de interés, como: la presa se encuentra casi vacía, casi llena o a media capacidad. Será de alcance limitado medir o predecir eventos de riesgo, como sequía o desborde.

3. Si $A, B \subset \Omega$, con $A \neq B$ y $A, B \neq \varnothing$, la σ-álgebra generada por los conjuntos A y B es
$$\begin{aligned}\mathcal{A} = \{\varnothing, \quad & A, \quad B, \quad A \cup B, \quad A \cap B, \quad A \cup B^c, \quad A \cap B^c, \\ \Omega, \quad & A^c, \quad B^c, \quad A^c \cap B^c, \quad A^c \cup B^c, \quad A^c \cap B, \quad A^c \cup B\}.\end{aligned}$$
Aún escasa, esta σ-álgebra es más rica que las dos anteriores.

4. La σ-álgebra *potencia* de Ω es
$$\mathcal{A} = 2^\Omega : \text{colección de los subconjuntos de } \Omega.$$
Esta es la σ-álgebra más grande del espacio muestral Ω, en el sentido de que contiene a cualquier otra σ-álgebra de Ω. La notación 2^Ω está inspirada en su cardinalidad. Si
$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\},$$
es un conjunto finito, con $\#\Omega = n \geq 1$, entonces habrá 2^n subconjuntos de Ω:
$$2^\Omega = \{\varnothing, \{\omega_1\}, \ldots, \{\omega_n\}, \{\omega_1, \omega_2\}, \ldots, \{\omega_{n-1}, \omega_n\}, \ldots, \Omega\}.$$

Algunas propiedades de las σ-álgebras son las siguientes.

- Si $\mathcal{A}_1 \subset \mathcal{A}_2$, son dos familias de subconjuntos de Ω, entonces, $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_2)$.

- Si $\mathcal{A}$ es una colección de conjuntos de Ω, entonces

$$\mathcal{A} \subset \sigma(\mathcal{A}) \subset 2^{\Omega}.$$

  Además, $\sigma(\mathcal{A}) = \mathcal{A}$ si y sólo si, $\mathcal{A}$ es una σ-álgebra.

- Si $\mathcal{A}$ y $\mathcal{B}$ son σ-álgebras de Ω, entonces $\mathcal{A} \cap \mathcal{B}$ es una σ-álgebra de Ω.

- Si $\{\mathcal{A}_\alpha\}_\alpha$ es una colección arbitraria de σ-álgebras de Ω, entonces $\bigcap_\alpha \mathcal{A}_\alpha$ es una σ-álgebra de Ω.

- Si $\mathcal{A}$ y $\mathcal{B}$ son σ-álgebras de Ω, entonces $\mathcal{A} \cup \mathcal{B}$ no necesariamente es una σ-álgebra de Ω. Sin embargo, recuerde que

$$\mathcal{A} \cup \mathcal{B} \subset \sigma(\mathcal{A} \cup \mathcal{B}).$$

- Si $\mathcal{A}$ es una colección de conjuntos de Ω, entonces su σ-álgebra generada es la intersección de todas las σ-álgebras en Ω que contienen a $\mathcal{A}$:

$$\sigma(\mathcal{A}) = \bigcap_{\substack{\mathcal{A} \subset \mathcal{F}, \\ \mathcal{F} \text{ es } \sigma\text{-álgebra}}} \mathcal{F}.$$

Ejemplo 1.2.8 (σ-álgebra de Borel)**.** Considere $\Omega = \mathbf{R}$ y $\mathcal{A}_1$ la colección de todos los intervalos abiertos:

$$\mathcal{A}_1 = \{(a,b) : a < b, a, b \in \mathbf{R}\}.$$

Esta familia de conjuntos no es una σ-álgebra en $\mathbf{R}$, ya que la unión de los intervalos abiertos $(1,3)$ y $(7,10)$ no pertenece a $\mathcal{A}_1$. Su σ-álgebra generada es la de *Borel*:

$$\mathcal{B}(\mathbf{R}) = \sigma(\mathcal{A}_1). \tag{1.7}$$

Esta es la mínima σ-álgebra en $\mathbf{R}$ que contiene a los intervalos abiertos. De hecho, $\mathcal{B}(\mathbf{R})$ contiene toda clase de intervalo: abierto, cerrado, semiabierto, semicerrado, semieje. Incluye además a sus complementos, uniones finitas o infinito numerables, intersecciones finitas o infinito numerables, entre otros. Se cumple también

$$\mathcal{B}(\mathbf{R}) = \sigma(\mathcal{A}_1) = \cdots = \sigma(\mathcal{A}_8), \tag{1.8}$$

donde

$$\begin{aligned}
\mathcal{A}_1 &= \{(a,b) : -\infty < a < b < \infty\}, & \mathcal{A}_5 &= \{(-\infty,b) : b \in \mathbf{R}\},\\
\mathcal{A}_2 &= \{[a,b] : -\infty < a < b < \infty\}, & \mathcal{A}_6 &= \{(-\infty,b] : b \in \mathbf{R}\},\\
\mathcal{A}_3 &= \{(a,b] : -\infty < a < b < \infty\}, & \mathcal{A}_7 &= \{(a,\infty) : a \in \mathbf{R}\},\\
\mathcal{A}_4 &= \{[a,b) : -\infty < a < b < \infty\}, & \mathcal{A}_8 &= \{[a,\infty) : a \in \mathbf{R}\}.
\end{aligned} \tag{1.9}$$

Por ejemplo, se verificará $\sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_6)$. Para ello, es suficiente comprobar

$$\mathcal{A}_1 \subset \sigma(\mathcal{A}_6) \quad \text{y} \quad \mathcal{A}_6 \subset \sigma(\mathcal{A}_1).$$

	espacio muestral Ω	cardinalidad $\#\Omega$	σ-álgebra usual
finito	$\{1,\ldots,n\}$ $\{\omega_1,\ldots,\omega_n\}$	n	2^Ω
infinito numerable	$\{0,1,\ldots\}$ $\{\omega_1,\omega_2,\ldots\}$	$\#\mathbf{N}$	2^Ω
infinito no numerable	$(0,1)$ $\mathbf{R}=(-\infty,\infty)$ $\mathbf{R}^2=\mathbf{R}\times\mathbf{R}$ $\mathbf{R}^n=\mathbf{R}\times\cdots\times\mathbf{R}$	$\#\mathbf{R}$	$\mathcal{B}((0,1))$ $\mathcal{B}(\mathbf{R})$ $\mathcal{B}(\mathbf{R}^2)$ $\mathcal{B}(\mathbf{R}^n)$

Tabla 1.6. Ejemplos de espacios muestrales, con su cardinalidad y σ-álgebra usual.

Sea el intervalo abierto $(a,b)\in\mathcal{A}_1$, con $-\infty<a<b<\infty$. Entonces

$$(a,b)=(a,\infty)\cap(-\infty,b)=((a,\infty)^c)^c\cap(-\infty,b)=(-\infty,a]^c\cap\bigcup_{n\geq 1}\left(-\infty,b-\frac{1}{n}\right]\in\sigma(\mathcal{A}_6).$$

Así, $\mathcal{A}_1\subset\sigma(\mathcal{A}_6)$. Recíprocamente, si $(-\infty,b]\in\mathcal{A}_6$, con $b\in\mathbf{R}$, entonces

$$(-\infty,b]^c=(b,\infty)=\bigcup_{n\geq 1}(b,b+n)\in\sigma(\mathcal{A}_1).$$

Así, $\mathcal{A}_6\subset\sigma(\mathcal{A}_1)$. Se concluye $\sigma(\mathcal{A}_1)=\sigma(\mathcal{A}_6)=\mathcal{B}(\mathbf{R})$. De manera similar, se verifica el resto de las igualdades de (1.8); véase Ejercicio 1.10.

Una importante pregunta es, ¿cuál σ-álgebra elegir de un espacio muestral dado Ω? Es deseable que la medida de probabilidad P se defina en una familia de conjuntos de Ω lo más rica posible. En este sentido, para un espacio muestral finito o infinito numerable, considere la σ-álgebra potencia 2^Ω. Sin embargo, la respuesta no es sencilla en un espacio muestral infinito no numerable. Su σ-álgebra potencia compromete el axioma aditivo numerable (1.11), descrita más adelante. Por lo cual, en el caso de $\Omega=\mathbf{R}$, se considera la σ-álgebra de Borel $\mathcal{B}(\mathbf{R})$. La Tabla 1.6 muestra ejemplos de espacios muestrales, sus respectivas cardinalidades, así como sus σ-álgebras usuales. En los ejemplos de dicha tabla, los elementos del espacio muestral Ω son eventos simples.

1.3. Espacio de probabilidad

En esta sección se construirán los espacios de probabilidad.

Un espacio de probabilidad es un modelo o estructura matemática que explica la naturaleza aleatoria de un fenómeno observado de la naturaleza o un experimento controlado. La construcción de un modelo de probabilidad en un experimento se resume con las siguientes tres etapas.

espacio muestral	Ω	conjunto de resultados posibles del experimento
σ-álgebra	$\mathcal{A}$	colección de eventos (subconjuntos de Ω)
medida de probabilidad	P	$0 \leq P(A) \leq 1$; para cada $A \in \mathcal{A}$ $(A \subset \Omega)$

Tabla 1.7. Elementos del modelo matemático espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

1. Establecer el *espacio muestral* Ω, que es el conjunto de los resultados posibles del experimento. El experimento puede ser controlado o bien basado en hechos observados de la naturaleza o de índole social.

2. Determinar la familia de eventos de interés $\mathcal{A}$, llamada σ-álgebra, donde será definida la media de probabilidad. La Tabla 1.6 describen ejemplos de las σ-álgebra usuales según la cardinalidad del espacio muestral Ω. Para un espacio muestral finito o infinito numerable, se sugiere la σ-álgebra potencia 2^{Ω}, de todos los subconjuntos de Ω. Para un espacio muestral euclidiano, como el intervalo $(0, 1)$ o la recta real $\mathbf{R}$, que son conjuntos infinito no numerables, se trabaja con su respectiva σ-álgebra de Borel.

3. Por último, se asigna la medida de probabilidad P a cada uno de los eventos de la σ-álgebra $\mathcal{A}$. Entre otras propiedades, esta medida satisface $0 \leq P(A) \leq 1$; para todo $A \in \mathcal{A}$.

La Tabla 1.7 describe los tres elementos de un espacio de probabilidad.

El siguiente ejemplo describe la construcción de un espacio de probabilidad como modelo matemático.

Ejemplo 1.3.1. Considere que se juega una vez a la ruleta; véase la Figura 1.4. Se describirá el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ de este experimento. El espacio muestral es el conjunto de los resultados posibles del experimento

$$\Omega = \{0, 00, 1, 2, \ldots, 36\}.$$

Su cardinalidad es finita: $\#\Omega = 38$. Como σ-álgebra, considere la colección de subconjuntos de Ω. El número de eventos que contiene esta σ-álgebra es $\#\mathcal{A} = 2^{38} = 274\,877\,906\,944 > 2.7 \times 10^{11}$. Ejemplos de eventos son

$$\varnothing, \Omega, \{0\}, \{00\}, \{1\}, \ldots, \{36\}, ROJO, NEGRO,$$

así como

$$\begin{aligned}
PAR &= \{2, \ldots, 36\}, \\
IMPAR &= \{1, \ldots, 35\}, \\
CHICO &= \{1, \ldots, 18\}, \\
GRANDE &= \{19, \ldots, 36\}, \\
FILA_1 &= \{1, 4, \ldots, 34\} = \{3i - 2 : i = 1, \ldots, 12\}, \\
FILA_2 &= \{2, 5, \ldots, 35\} = \{3i - 1 : i = 1, \ldots, 12\}, \\
FILA_3 &= \{3, 6, \ldots, 36\} = \{3i : i = 1, \ldots, 12\}.
\end{aligned}$$

Eventos son también la unión o intersección finita de cualesquiera de ellos, junto con sus complementos. Los eventos PAR e $IMPAR$ son excluyentes, aunque no son el complemento uno del otro:

$$PAR^c = IMPAR \cup \{0, 00\},$$
$$IMPAR^c = PAR \cup \{0, 00\}.$$

Por último, la probabilidad asignada a cada uno de los eventos de Ω es

$$P(A) = \frac{\#A}{\#\Omega}; \quad \text{para} \quad A \in \mathcal{A} \quad (A \subset \Omega). \tag{1.10}$$

Esta fórmula aplica sólo cunado los eventos simples tienen la misma probabilidad (equiprobables). En este ejemplo se tiene

$$P(0) = P(00) = P(1) \cdots = P(36) = \frac{\#\{0\}}{\#\Omega} = \frac{1}{38} = 0.0263 = 2.63\,\%.$$

Al aplicar (1.10), se obtiene

$$P(\Omega) = \frac{\#\Omega}{\#\Omega} = \frac{38}{38} = 1 = 100\,\%,$$
$$P(ROJO) = \frac{\#ROJO}{\#\Omega} = \frac{18}{38} = 0.4737 = 47.37\,\%,$$
$$P(NEGRO) = \frac{\#NEGRO}{\#\Omega} = 0.4737,$$
$$P(CHICO) = \frac{\#CHICO}{\#\Omega} = 0.4737,$$
$$P(FILA_1) = \frac{\#FILA_1}{\#\Omega} = \frac{12}{38} = 0.3158.$$

Cabe mencionar que no aplica la fórmula (1.10) cuando los eventos simples no son equiprobables. Por ejemplo, considere el experimento de lanzar un dado y registrar al caer el número de la cara superior. El resultado del experimento es uno de los siguientes eventos simples: $1, 2, 3, 4, 5, 6$. Es natural asumir que el dado es simétrico. En este sentido, los eventos simples son equiprobables

$$P(1) = \cdots = P(6) = \frac{1}{6} = 0.1667 = 16.67\,\%.$$

Sin embargo, si el dado está cargado o asimétrico, ya sea por fallas de diseño o producción, desgaste o alguna otra causa, los eventos simples tienen diferentes probabilidades. En tal caso, a cada evento simple se le asigna una probabilidad genérica o a especificar:

$$P(1) = p_1, \qquad P(2) = p_2, \qquad P(3) = p_3,$$
$$P(4) = p_4, \qquad P(5) = p_5, \qquad P(6) = p_6,$$

donde

$$p_1, \ldots, p_6 \geq 0 \quad \text{y} \quad p_1 + \cdots + p_6 = 1.$$

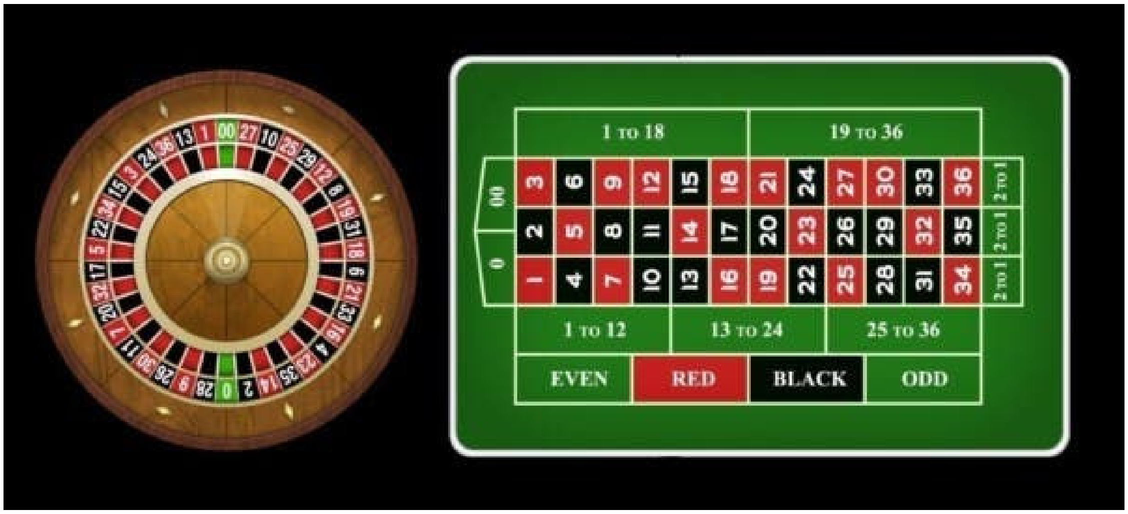


Figura 1.4. Ruleta de 38 números; Ejemplo 1.3.1.

Dichas probabilidades son fijas, no necesariamente iguales y tal vez desconocidas.

La probabilidad interviene en la modelación de la incertidumbre de experimentos con eventos más complejos. Imagine por ejemplo a una planta petrolera, cuya preocupación es evitar accidentes. Entre otros factores, los eventos de fallas dependen de la calidad de los componentes de la planta, como motores, tubería, estructuras metálicas, estado de salud y profesionalismo del personal. Por lo tanto se debe proponer un modelo matemático que tome en cuenta todas, o al menos las más significativas, las fuente de variabilidad que perturban el estado de la calidad de la planta.

Ejemplo 1.3.2 (Continuación de Ejemplo 1.2.4)**.** Considere el experimento de lanzar una moneda tres veces y registrar la cara superior. En cada lanzamiento, los resultados posibles son $a =$ águila ó $s =$ sol. El espacio muestral es

$$\Omega = \{xyz : x, y, z = a, s\} = \{aaa, aas, asa, ass, saa, sas, ssa, sss\}.$$

Al ser un espacio muestral finito, se considerará como σ-álgebra a la colección de todos los subconjuntos de Ω; véase la Tabla 1.6. Como $\#\Omega = 8$, entonces hay $2^8 = 256$ subconjuntos de Ω; véase Ejercicio 1.11. La asignación de probabilidad de cada uno de estos eventos se determina por la de los eventos simples. Se se asume que la moneda es legal o simétrica, todos los eventos simples son equiprobables; tienen la misma probabilidad de ocurrir:

$$P(aaa) = P(aas) = \cdots = P(sss) = \frac{1}{\#\Omega} = \frac{1}{8} = 0.125 = 12.5\,\%.$$

Así mismo, para los eventos

$$\begin{aligned} &A : \text{hay al menos un águila} = \{aaa, aas, asa, ass, saa, sas, ssa\}, \\ &B : \text{hay exactamente dos soles} = \{ass, sas, ssa\}, \end{aligned}$$

se tiene

$$P(A) = \frac{\#A}{\#\Omega} = \frac{7}{8} = 0.875 = 87.5\,\%,$$
$$P(B) = \frac{\#B}{\#\Omega} = \frac{3}{8} = 0.375 = 37.5\,\%.$$

El evento A es muy probable de ocurrir mientras que la probabilidad del evento B es baja. De manera análoga, la probabilidad se define en todos los eventos de la σ-álgebra; véase el Ejercicio 1.11. Por último, cabe mencionar que hay variantes del modelo que considera iguales algunos eventos simples, como *aas* y *asa*. Esto se debe a que, en el registro de los resultados, no importa el orden de la muestra. En ese caso, se reduce el espacio muestral y cambian las probabilidades de los eventos simples.

Definición 1.3.3. Considere un experimento con espacio muestral no vacío Ω y una σ-álgebra $\mathcal{A}$. Se dice que una función real P, definida en $\mathcal{A}$, es una *medida de probabilidad* si satisface los *axiomas de Kolmogorov*:

1. $$P(\Omega) = 1.$$

2. Para todo $A \in \mathcal{A}$:
$$P(A) \geq 0.$$

3. Para cualquier colección numerable y excluyente de conjuntos $A_1, A_2, \dots \in \mathcal{A}$, se tiene
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n). \tag{1.11}$$

A la terna $(\Omega, \mathcal{A}, P)$ se le llama *espacio de probabilidad*. Se dice que A es un *evento* de Ω si pertenece a $\mathcal{A}$. En particular $\omega \in \Omega$ es un *evento simple*, siempre que $\{\omega\} \in \mathcal{A}$.

Los axiomas de Kolmogorov son intuitivos. El primero da estatus de evento *seguro* al espacio muestral Ω. De modo que siempre ocurrirá. El segundo axioma otorga una probabilidad no negativa a todos los eventos. Por último, el tercer axioma es la propiedad *aditiva numerable* de la medida P. Propiedad que es análoga al cálculo del área de una unión de superficies excluyentes.

Ejemplo 1.3.4. Los siguientes son ejemplos de espacio de probabilidad.

1. **Espacio de probabilidad binario (éxito-fracaso)**. Este modelo aplica al fenómeno de incertidumbre en donde sólo hay dos resultados posibles: éxito ó fracaso. Este es el espacio de probabilidad práctico más sencillo posible. Lo que contrasta con su relevancia de aplicación en todas las áreas del conocimiento; en redes neuronales, por ejemplo. El espacio muestral es
$$\Omega = \{\text{fracaso}, \text{éxito}\} = \{0, 1\}.$$
La σ-álgebra potencia se compone de $2^2 = 4$ eventos:
$$\mathcal{A} = \{\varnothing, \text{fracaso}, \text{éxito}, \Omega\}.$$

éxito	fracaso	$p = P(\text{éxito})$
águila	sol	0.5
sol	águila	0.5
rojo en ruleta	no rojo	0.4737
00 en ruleta	no 00	0.0263
13 en ruleta	no 13	0.0263
llegar temprano a cita	dilatar	$p \approx 0.7$
vida de pantalla > 1 año	dura un año o menos	$p \approx 1$
vida de celular ≤ 1 año	dura más de un año	$p \approx 0$
fármaco alivia a paciente	no lo alivia	0.65

Tabla 1.8. Ejemplos de espacio de probabilidad binario.

Las probabilidades de los dos únicos eventos simples son

$$P(\text{éxito}) = P(1) = p,$$
$$P(\text{fracaso}) = P(0) = 1 - p, \quad \text{con} \quad 0 \leq p \leq 1.$$

Algunos ejemplos de éxito y fracaso se describen en la Tabla 1.8. En el lanzamiento de una moneda, éxito es águila y fracaso sol. La asignación de éxito y fracaso puede ser al revés, si así se considera conveniente. Note que algunos de estos ejemplos son simplificaciones de eventos de espacios muestrales más complejos. En el caso de la ruleta, el evento que sea rojo se considera simple. En cambio, según el espacio muestral descrito en el Ejemplo 1.3.1, dicho evento es una unión de 18 eventos simples. En cualquier caso, la probabilidad es la misma: $P(ROJO) = 18/38$.

2. **Espacio de probabilidad finito**. Aquí el espacio muestral Ω se compone de un número finito eventos simples. Ejemplos de este tipo de espacios se encuentran los que se han descrito hasta el momento: el de lanzamiento de una moneda o un dado, una o varias veces. Así también se encuentran los ejemplos de la ruleta y el del espacio de probabilidad binario 1.3.4.1.

   Considere el espacio muestral finito $\Omega = \{\omega_1, \ldots, \omega_n\}$, con $n > 1$. Como σ-álgebra para Ω, tome su σ-álgebra potencia $\mathcal{A} = 2^\Omega$. El siguiente paso consiste en definir la probabilidad en cada uno de los eventos simples:

$$P(\omega_1) = p_1, \quad P(\omega_2) = p_2, \ldots, \quad P(\omega_n) = p_n;$$

   con

$$p_1, p_2, \ldots, p_n \geq 0 \quad \text{y} \quad p_1 + p_2 + \cdots + p_n = 1. \tag{1.12}$$

   Para un evento $A \in \mathcal{A}$ $(A \subset \Omega)$, con $A = \{\omega_{n_1}, \omega_{n_2}, \ldots, \omega_{n_k}\}$, para ciertos $1 \leq n_1 < n_2 <$

$\cdots < n_k \leq n$ y $1 \leq k \leq n$. La medida P se define como

$$\begin{aligned} P(A) &= P(\{\omega_{n_1}, \omega_{n_2}, \ldots, \omega_{n_k}\}) \\ &= P(\omega_{n_1}) + P(\omega_{n_2}) + \cdots + P(\omega_{n_k}) \\ &= p_{n_1} + p_{n_2} + \cdots + p_{n_k}. \end{aligned}$$

La verificación de los axiomas de Kolmogorov resulta inmediata. 1) Note que

$$P(\Omega) = P(\{\omega_1, \ldots, \omega_n\}) = p_1 + p_2 + \cdots + p_n = 1.$$

2) El segundo axioma también se cumple:

$$P(A) = \sum_{i=1}^{k} p_{n_i} \geq 0.$$

3) Ahora bien, como Ω tiene un número finito de elementos entonces, $\mathcal{A}$ tiene también un número finito de eventos. Así que toda colección excluyente de eventos debe ser finita. Así, considere a $A_1, \ldots, A_k \in \mathcal{A}$, una colección excluyente de eventos de modo que

$$A = \bigcup_{i=1}^{k} A_i = \bigcup_{i=1}^{k} \bigcup_{\omega_{ij} \in A_i} \{\omega_{ij}\} = \bigcup_{\omega_i \in A} \{\omega_i\}.$$

Todos los eventos simples de la expresión anterior son diferentes entre sí. Por lo que

$$P(A) = P\left(\bigcup_{\omega_i \in A} \{\omega_i\}\right) = \sum_{\omega_i \in A} p_i = \sum_{i=1}^{k} \sum_{\omega_{ij} \in A_i} p_{ij} = \sum_{i=1}^{k} P(A_i).$$

Se concluye que P es una media de probabilidad definida en $\mathcal{A}$ y $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad. Note que hay muchas medidas de probabilidad que se pueden definir en la σ-álgebra $\mathcal{A}$. Tan solo hay que cuidar que la suma de las probabilidades de los eventos simples sea uno: (1.12).

Si en particular los eventos simples son equiprobables, entonces

$$p_1 = p_2 = \cdots = p_n = \frac{1}{n}$$

con $n = \#\Omega$. En consecuencia, la probabilidad de un evento A se determina por el número de eventos simples que contiene:

$$P(A) = \frac{\text{número de eventos simples de } A}{\text{número de eventos simples de } \Omega} = \frac{\#A}{\#\Omega}; \quad A \subset \Omega. \tag{1.13}$$

3. **Espacio de probabilidad discreto**. Un espacio muestral no vacío Ω es discreto si es finito o infinito numerable. El caso finito se analizó en el inciso anterior. Para el caso infinito numerable, considere que el espacio muestral es una sucesión infinita numerable de eventos simples: $\Omega = \{\omega_1, \omega_2, \ldots\}$, junto con su σ-álgebra potencia $\mathcal{A} = 2^{\Omega}$. Como

en el caso del espacio de probabilidad finito, la media de probabilidad P se determina al definirse en cada uno de los eventos simples:

$$P(\omega_1) = p_1, \quad P(\omega_2) = p_2, \dots,$$

bajo las restricciones

$$p_i \geq 0; \quad \text{para} \quad i \geq 1 \quad \text{y} \quad \sum_{i=1}^{\infty} p_i = 1.$$

Así, la medida de probabilidad P se define en cualquier subconjunto del espacio muestral como sigue:

$$P(A) = \sum_{\omega_i \in A} p_i; \quad A \subset \Omega.$$

Por ejemplo, considere el experimento de observar el número de goles de un equipo de fútbol en su próximo encuentro. El espacio muestral es $\Omega = \{0, 1, 2, \dots\}$. Asigne las siguientes probabilidades a cada uno de los eventos simples

$$\begin{aligned}
P(0) &= p_0 = 0.2, \\
P(1) &= p_1 = 0.2 \times 0.8 = 0.16, \\
P(2) &= p_2 = 0.2 \times 0.8^2 = 0.128, \\
P(3) &= p_3 = 0.2 \times 0.8^3 = 0.1024, \\
&\vdots \\
P(i) &= p_i = 0.2 \times 0.8^i, \\
&\vdots
\end{aligned}$$

La suma de estas probabilidades es uno:

$$\begin{aligned}
\sum_{i=0}^{\infty} p_i &= \sum_{i=0}^{\infty} 0.2 \times 0.8^i = 0.2(0.8^0 + 0.8^1 + 0.8^2 + 0.8^3 + \cdots) \\
&= 0.2(1 + a + a^2 + a^3 + \cdots) = 0.2 \times \frac{1}{1-a} \\
&= \frac{0.2}{1-0.8} = \frac{0.2}{0.2} \\
&= 1.
\end{aligned}$$

Aquí se usó la serie geométrica (22) del Apéndice, con $a = 0.8$.

4. **Espacio de probabilidad continuo**. Este tipo de espacios se asocia a las variables aleatorias continuas, en donde la probabilidad se define como el área bajo la curva de una función real positiva dada; véase el Capítulo 3, más adelante. Sea $\Omega = \mathbf{R}$, con $\mathcal{A} = \mathcal{B}(\mathbf{R})$, su σ-álgebra de Borel. Recuerde que $\mathcal{B}(\mathbf{R})$ contiene toda clase de intervalo:

$$(a,b), [a,b], (a,b], [a,b), (-\infty,b), (-\infty,b], (a,\infty), [a,\infty); \quad \text{con} \quad -\infty < a \leq b < \infty.$$

El conjunto vacío $\varnothing$ y la recta real $\mathbf{R} = (-\infty, \infty)$ pertenecen también a $\mathcal{B}(\mathbf{R})$. De hecho ambos son intervalos, tanto abiertos como cerrados. Considere una función real integrable $f(x)$ tal que

$$f(x) \geq 0 \quad \text{y} \quad \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1. \tag{1.14}$$

Considere la función P evaluada en el intervalo semiabierto $(a, b]$ como

$$P((a, b]) = \int_a^b f(t)\,\mathrm{d}t\,. \tag{1.15}$$

Esta función induce una única medida de probabilidad en $\mathcal{B}(\mathbf{R})$, tal que satisfaga (1.15). Por propiedades de la integral, se tiene

$$P((a, b)) = P([a, b]) = P(([a, b)) = P((a, b]) = \int_a^b f(t)\,\mathrm{d}t\,.$$

Así mismo

$$P((-\infty, b)) = P((-\infty, b]) = \int_{-\infty}^b f(t)\,\mathrm{d}t\,,$$

$$P((a, \infty)) = P([a, \infty)) = \int_a^{\infty} f(t)\,\mathrm{d}t$$

y

$$P(\mathbf{R}) = P((-\infty, \infty)) = \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1.$$

Un ejemplo de espacio de probabilidad continuo es la distribución exponencial de tasa uno:

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

Por el Ejercicio 1.17, la función $f(x)$ satisface las restricciones (1.14). Véase también la Sección 3.3, más adelante.

La siguiente proposición muestra las propiedades básicas de la probabilidad, que se deducen de los axiomas de Kolmogorov.

Proposición 1.3.5. Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y A, B dos eventos en $\mathcal{A}$. Entonces

1. $$P(A^c) = 1 - P(A).$$

   En particular

   $$P(\varnothing) = 0.$$

2. $$P(A \cup B) = P(A) + P(B) - P(A \cap B), \tag{1.16}$$
   $$P(A \cap B) = P(A) + P(B) - P(A \cup B). \tag{1.17}$$

3. *Ley de probabilidad total*

$$P(A) = P(A \cap B) + P(A \cap B^c). \tag{1.18}$$

Más general, si $\{B_n\}_{n\geq 1}$ es una partición de Ω, entonces

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n). \tag{1.19}$$

Primero se deducirá la ley de probabilidad total (1.19), pues con esta se obtienen las demás propiedades. Recuerde que $\{B_n\}_{n\geq 1}$ es una partición de Ω si es una colección excluyente de eventos cuya unión resulta Ω:

$$B_i \cap B_j = \varnothing; \quad \text{para} \quad i, j \geq 1,\ i \neq j$$

y

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

Demostración. 3. La partición $\{B_n\}_{n\geq 1}$ de Ω, induce también una partición del evento A:

$$A = A \cap \Omega = A \cap \left(\bigcup_{n=1}^{\infty} B_n \right) = \bigcup_{n=1}^{\infty} (A \cap B_n),$$

donde $\{A \cap B_n\}_{n\geq 1}$, es una colección de eventos excluyentes. Por el tercer axioma de Kolmogorov (1.11), se tiene

$$P(A) = P\left(\bigcup_{n=1}^{\infty} (A \cap B_n) \right) = \sum_{n=1}^{\infty} P(A \cap B_n).$$

En particular, si B es cualquier evento, entonces B y B^c crean una partición de Ω, así como inducen una partición de A:

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$$

y

$$P(A) = P((A \cap B) \cup (A \cap B^c)) = P(A \cap B) + P(A \cap B^c). \tag{1.20}$$

1. Note que A y A^c forman una partición de Ω:

$$A \cup A^c = \Omega \quad \text{y} \quad A \cap A^c = \varnothing.$$

Por lo que

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

En particular como $\varnothing^c = \Omega$, entonces

$$P(\varnothing) = 1 - P(\varnothing^c) = 1 - P(\Omega) = 1 - 1 = 0.$$

2. Aquí considere dos particiones simultaneas:

$$
\begin{aligned}
A &= (A \cap B) \cup (A \cap B^c), \\
B &= (A \cap B) \cup (A^c \cap B).
\end{aligned}
$$

Así

$$
\begin{aligned}
A \cup B &= ((A \cap B) \cup (A \cap B^c)) \cup ((A \cap B) \cup (A^c \cap B)) \\
&= (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B).
\end{aligned}
$$

La última unión es de tres eventos excluyentes. Por lo que

$$
\begin{aligned}
P(A) &= P(A \cap B) + P(A \cap B^c), \\
P(B) &= P(A \cap B) + P(A^c \cap B), \\
P(A \cup B) &= P(A \cap B) + P(A \cap B^c) + P(A^c \cap B).
\end{aligned}
$$

De esta manera

$$
\begin{aligned}
P(A \cup B) &= P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) \\
&= P(A \cap B) + [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] \\
&= P(A) + P(B) - P(A \cap B).
\end{aligned}
$$

$\square$

La siguiente proposición resume las propiedades básicas de la probabilidad en el contexto de desigualdades.

Proposición 1.3.6. Sean los eventos de un espacio muestral Ω: A, B y $A_1, A_2, \ldots$.

1. Si $A \subset B$, entonces
$$P(A) \leq P(B).$$
El recíproco no es cierto. En particular
$$0 \leq P(A) \leq 1 \tag{1.21}$$
y
$$P(A \cap B) \leq P(A) \leq P(A \cup B).$$

2.
$$P(A \cup B) \leq P(A) + P(B).$$
En general
$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n). \tag{1.22}$$

Demostración. 1. Aquí construya una partición del evento B en términos de A y su complemento A^c:

$$B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c) = A \cup (B \cap A^c).$$

La última unión es de eventos excluyentes. Por lo que

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

En particular, la segunda desigualdad en (1.21), se obtiene al considerar la relación $A \subset \Omega$. Lo que implica

$$P(A) \leq P(\Omega) = 1.$$

Por último, como

$$A \cap B \subset A \subset A \cup B,$$

entonces

$$P(A \cap B) < P(A) \leq P(A \cup B).$$

2. Para verificar la propiedad (1.22), exprese $\bigcup_{n=1}^{\infty} A_n$ como una unión de eventos excluyentes:

$$\begin{array}{ccccccccc} & A_1 & \cup & A_2 & \cup & A_3 & \cup & \cdots \\ = & A_1 & \cup & (A_2 \cap A_1^c) & \cup & (A_3 \cap [A_1 \cup A_2]^c) & \cup & \cdots \\ = & A_1 & \cup & (A_2 \cap A_1^c) & \cup & (A_3 \cap A_1^c \cap A_2^c) & \cup & \cdots \end{array}$$

Por lo que

$$\begin{array}{ll} & P(A_1 \cup A_2 \cup A_3 \cup \cdots) \\ = & P(A_1 \cup (A_2 \cap A_1^c) \cup (A_3 \cap A_1^c \cap A_2^c) \cup \cdots) \\ = & P(A_1) \quad + \quad P(A_2 \cap A_1^c) \quad + \quad P(A_3 \cap A_1 \cap A_2^c) \quad + \quad \cdots \\ \leq & P(A_1) \quad + \quad P(A_2) \quad + \quad P(A_3) \quad + \quad \cdots \end{array}$$

□

La siguiente proposición describe la propiedad de continuidad de la probabilidad, de modo que el límite de la probabilidad de una sucesión creciente o decreciente de eventos es la probabilidad del límite de dichos eventos. Además, tanto (1.23) como (1.24), son propiedades equivalentes al axioma aditivo numerable (1.11); véase el Ejercicio 1.16.

Proposición 1.3.7 (continuidad de la probabilidad)**.** Sea $\{A_n\}_{n\geq 1}$ una sucesión de eventos en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.
1. Si $\{A_n\}_{n\geq 1}$ es creciente:

$$A_n \uparrow A; \quad n \to \infty, \quad \text{con} \quad A = \bigcup_{n=1}^{\infty} A_n,$$

entonces

$$\lim_{n\to\infty} P(A_n) = P(A). \tag{1.23}$$

2. Si $\{A_n\}_{n\geq 1}$ es decreciente:

$$A_n \downarrow A; \quad n \to \infty, \quad \text{con} \quad A = \bigcap_{n=1}^{\infty} A_n,$$

entonces

$$\lim_{n\to\infty} P(A_n) = P(A). \tag{1.24}$$

Demostración. 1. Como

$$A_1 \subset A_2 \subset \cdots \subset A = \bigcup_{n=1}^{\infty} A_n,$$

entonces

$$P(A_1) \leq P(A_2) \leq \cdots \leq P(A).$$

La sucesión de probabilidades $\{P(A_n)\}_{n=1}^{\infty}$, es creciente y acotada superiormente por $P(A)$. Por lo que es convergente. Con la técnica de las "donas", se probará que $P(A)$ es su límite, es decir, su supremo. Sea la sucesión de eventos excluyentes $\{D_n\}_{n\geq 1}$, definida como

$$D_n = A_n \cap A_{n-1}^c; \quad n = 1, 2, \ldots,$$

con $A_0 = \varnothing$. Luego

$$\begin{aligned} A_n &= A_1 \cup A_2 \cup \cdots \cup A_n \\ &= A_1 \cup (A_2 \cap A_1^c) \cup \cdots \cup (A_n \cap A_{n-1}^c) \\ &= D_1 \cup D_2 \cup \cdots \cup D_n. \end{aligned}$$

Por lo que

$$A_n = \bigcup_{i=1}^{n} D_i$$

y

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left(\bigcup_{i=1}^{n} D_i \right) = \bigcup_{n=1}^{\infty} D_n = A.$$

Luego

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

y

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(D_i) = \sum_{n=1}^{\infty} P(D_n) = P\left(\bigcup_{n=1}^{\infty} D_n \right) = P(A).$$

2. Como $A_n \downarrow A$; $n \to \infty$, con $A = \cap_{n\geq 1} A_n$, entonces $A_n^c \uparrow A^c$; $n \to \infty$. En efecto

$$A_1 \supset A_2 \supset \cdots \supset A,$$

implica

$$A_1^c \subset A_2^c \subset \cdots \subset A^c$$

y

$$A^c = \left(\bigcap_{n \geq 1} A_n \right)^c = \bigcup_{n \geq 1} A_n^c.$$

La última desigualdad se debe a las leyes de De Morgan; véanse Sección 1.2.1 y Ejercicio 1.15. Del caso creciente, se tiene

$$\lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} [1 - P(A_n^c)] = 1 - \lim_{n \to \infty} P(A_n^c) = 1 - P(A^c) = P(A).$$

$\square$

1.4. Muestreo de poblaciones finitas y técnicas de conteo

En esta sección se describirán los esquemas básicos de muestreo para poblaciones finitas, así como el muestreo hipergeométrico.

Un experimento controlado implica un diseño o esquema de muestreo,que a su vez induce un modelo de probabilidad. Las técnicas de conteo son útiles tanto para obtener la cardinalidad de los eventos de interés como del espacio muestral. En ese caso el espacio de probabilidad es finito. De esta manera se calcula la probabilidad de dichos eventos.

En primera instancia, se abordará el espacio de probabilidad uniforme, surgidos de diversos esquemas de muestreo. De esta manera, la probabilidad de un evento $A \subset \Omega$, es proporcional al total de eventos simples que contiene. Por expresión (1.13), se tiene

$$P(A) = \frac{\text{total de eventos simples de } A}{\text{total de eventos simples de } \Omega} = \frac{\#A}{\#\Omega}; \quad A \subset \Omega.$$

Con esta fórmula, el cálculo de la probabilidad se reduce a contar el número de elementos del evento A, así como del espacio muestral Ω.

Las técnicas de conteo se apoyan en el teorema fundamental de conteo, que se enuncia a continuación.

Teorema 1.4.1 (Teorema fundamental de conteo)**.** Sea un trabajo que consiste de k tareas, tal que la i-ésima tarea se realiza de n_i diferentes maneras; para $i = 1, \ldots, k$. Entonces, el trabajo se realiza de $n_1 \times n_2 \times \cdots \times n_k$ diferentes maneras.

La siguiente figura representa el conteo de k tareas en un arreglo de k casilleros o elementos ordenados.

$$\begin{array}{cccc} n_1 & n_2 & \cdots & n_k \\ \downarrow & \downarrow & & \downarrow \\ \boxed{} & \boxed{} & \cdots & \boxed{} \end{array}$$

La primera tarea se realiza de n_1 maneras diferentes. Cada una de ellas se acopla con las n_2 formas de llevar a cabo la segunda tarea. Hasta aquí se tienen $n_1 \times n_2$ diferentes maneras de realizar las tareas primera y segunda. Al continuar sucesivamente, el total de maneras de realizar el trabajo es $n_1 \times n_2 \times \cdots \times n_k$.

Ejemplo 1.4.2. Se muestran algunas aplicaciones del teorema fundamental de conteo.

1. Se lanza un dado $k = 2$ veces y registra el número de su cara superior, como lo describe el siguiente arreglo de dos dimensiones:

$$\begin{array}{|c|c|} \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} \\ \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} \\ \hline & \\ \hline \end{array}$$

   El espacio muestral es $\Omega = \{(x_1, x_2) : x_i = 1, 2, 3, 4, 5, 6;\ i = 1, 2\}$. Mientras que su cardinalidad resulta
$$\#\Omega = 6 \times 6 = 6^2 = 36.$$
   Recuerde la representación gráfica del espacio muestral Ω en la Tabla 1.5.

2. En cambio, al lanzar el dado $k = 5$ veces, el espacio muestral de este otro experimento es $\Omega = \{(x_1, x_2, x_3, x_4, x_5) : x_i = 1, 2, 3, 4, 5, 6;\ i = 1, 2, 3, 4, 5\}$. Su cardinalidad es
$$\#\Omega = 6 \times 6 \times 6 \times 6 \times 6 = 6^5 = 7\,776.$$
   El siguiente arreglo representa a los eventos simples de Ω.

$$\begin{array}{|c|c|c|c|c|} \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} \\ \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} \\ \hline & & & & \\ \hline \end{array}$$

3. Se lanza un dado 2 veces y una moneda 3 veces. El espacio muestral de todos los resultados posibles del experimento es
$$\Omega = \{(x_1, x_2, x_3, x_4, x_5) : x_1, x_2 = 1, \ldots, 6,\ \ x_3, x_4, x_5 = 0, 1\},$$
   El arreglo correspondiente se representa en la siguiente figura.

$$\begin{array}{|c|c|c|c|c|} \multicolumn{1}{c}{6} & \multicolumn{1}{c}{6} & \multicolumn{1}{c}{2} & \multicolumn{1}{c}{2} & \multicolumn{1}{c}{2} \\ \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} \\ \hline & & & & \\ \hline \end{array}$$

   Así
$$\#\Omega = 6^2 \times 2^3 = 288.$$

4. El sorteo de la *Lotería Nacional* tiene una emisión de 50000 boletos, numerados del 0 al 49999. Para el sorteo se usan 5 tómbolas. La primera contiene 5 bolas, numeradas como $0, 1, 2, 3, 4$. Las otras cuatro tómbolas contienen 10 bolas cada una, numeradas de 0 al 9. Los arreglos posibles se describen en la siguiente figura.

$$\begin{array}{ccccc} 5 & 10 & 10 & 10 & 10 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \end{array}$$

La cardinalidad del espacio muestral es

$$\#\Omega = 5 \times 10^4 = 50000.$$

5. Las placas de automóvil en México tienen un formato de $k = 9$ campos, que se llenan con los siguientes $n = 38$ caracteres: alfabeto latino de 26 letras, dígitos del 0 al 9, guión "-" y espacio hueco "☐". Este último carácter está bien definido y no debe confundirse con el conjunto vacío $\varnothing$. Los caracteres representan la población mientras que los campos son la muestra, como se describe en la siguiente arreglo.

$$\begin{array}{ccccccccc} 38 & 38 & 38 & 38 & 38 & 38 & 38 & 38 & 38 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \end{array}$$

Se selecciona al azar un carácter de los 38 posibles, cuyo resultado se registra en el primer casillero. Luego, se selecciona al azar otro carácter, que puede ser el mismo de la primera extracción. Lo que representa otros 38 resultados posibles. Hasta aquí hay 38×38 diferentes resultados posibles para los primeros dos campos. Se continua de la misma manera hasta obtener $k = 9$ extracciones con reemplazo. El total de placas posibles es más de cien billones:

$$\#\Omega = 38^9 = 1.65 \times 10^{14}.$$

Cabe mencionar que esta capacidad rebasa con mucho al parque vehicular actual de México. Según datos del INEGI [8], al final del año 2021 había $53,115,396$ millones de vehículos de motor.

6. Una *baraja* francesa (inglesa o americana) se compone de 52 cartas o naipes. Cada *carta* combina un número y una figura. Los números son

$$A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K,$$

mientras que las figuras resultan

$$\heartsuit \text{ corazón}, \ \clubsuit \text{ trébol}, \ \diamondsuit \text{ diamante}, \ \spadesuit \text{ espada o pica}.$$

La Tabla 1.9 describe el espacio muestral Ω; si se extrae al azar una carta. Su cardinalidad es $\#\Omega = 13 \times 4 = 52$, pues cada número se acopla con cada figura:

$$\begin{array}{cc} 13 & 4 \\ \downarrow & \downarrow \\ \boxed{\text{número}} & \boxed{\text{figura}} \end{array}$$

$A\heartsuit$	$2\heartsuit$	$3\heartsuit$	$4\heartsuit$	$5\heartsuit$	$6\heartsuit$	$7\heartsuit$	$8\heartsuit$	$9\heartsuit$	$10\heartsuit$	$J\heartsuit$	$Q\heartsuit$	$K\heartsuit$
$A\clubsuit$	$2\clubsuit$	$3\clubsuit$	$4\clubsuit$	$5\clubsuit$	$6\clubsuit$	$7\clubsuit$	$8\clubsuit$	$9\clubsuit$	$10\clubsuit$	$J\clubsuit$	$Q\clubsuit$	$K\clubsuit$
$A\diamondsuit$	$2\diamondsuit$	$3\diamondsuit$	$4\diamondsuit$	$5\diamondsuit$	$6\diamondsuit$	$7\diamondsuit$	$8\diamondsuit$	$9\diamondsuit$	$10\diamondsuit$	$J\diamondsuit$	$Q\diamondsuit$	$K\diamondsuit$
$A\spadesuit$	$2\spadesuit$	$3\spadesuit$	$4\spadesuit$	$5\spadesuit$	$6\spadesuit$	$7\spadesuit$	$8\spadesuit$	$9\spadesuit$	$10\spadesuit$	$J\spadesuit$	$Q\spadesuit$	$K\spadesuit$

Tabla 1.9. Espacio muestral del experimento de seleccionar al azar una carta; Ejemplo 1.4.2.6.

7. *Diseño de experimentos: modelo de tres factores*. En un experimento agrícola se desea medir la eficiencia de una semilla. Se considera que la productividad depende de tres *factores* o variables control: tipo de fertilizante, tipo de riego y temperatura media. Los factores tienen respectivamente a, b y c niveles. Entonces, el total de *tratamientos* o *celdas* es abc, según se describe en la siguiente figura.

$$\begin{array}{ccc} a & b & c \\ \downarrow & \downarrow & \downarrow \\ \boxed{} & \boxed{} & \boxed{} \end{array}$$

Una representación explícita de las abc celdas se muestra en la siguiente figura.

111	121	$\cdots$	$1b1$
211	221	$\cdots$	$2b1$
$\vdots$	$\vdots$		$\vdots$
$a11$	$a21$	$\cdots$	$ab1$

112	122	$\cdots$	$1b2$
212	222	$\cdots$	$2b2$
$\vdots$	$\vdots$		$\vdots$
$a12$	$a22$	$\cdots$	$ab2$

$\cdots$

$11c$	$12c$	$\cdots$	$1bc$
$21c$	$22c$	$\cdots$	$2bc$
$\vdots$	$\vdots$		$\vdots$
$a1c$	$a2c$	$\cdots$	abc

Ahora se describirán los cuatro esquemas básicos de muestreo para poblaciones finitas. El esquema de muestreo se determina según si la muestra es ordenada o no, así como, si hay o no reemplazo. La Tabla 1.10 describe los cuatro tipos de muestreo, dos de ellos denominados como *permutación* y *combinación*.

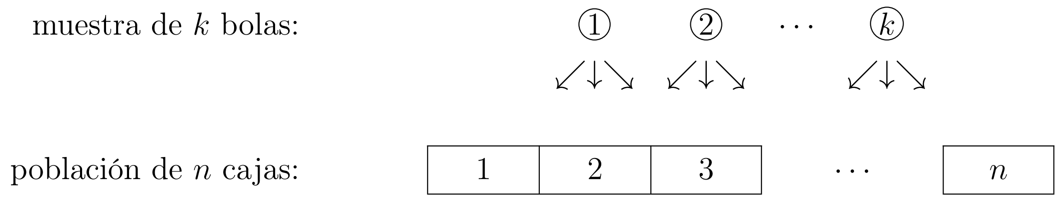
	con reemplazo	sin reemplazo
ordenado		permutación
no ordenado		combinación

Tabla 1.10. Tipos básicos de muestreo para poblaciones finitas.

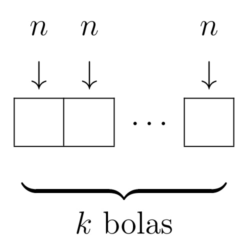
En lo que resta de esta sección, considere el experimento de seleccionar una muestra de tamaño $k \geq 1$ en una población finita de tamaño $n \geq 1$. Cada muestra representa un resultado posible (evento simple) del espacio muestral Ω. Su cardinalidad es finita, pues lo es la población origen. Cabe no confundir a población con espacio muestral.

1.4.1. Muestreo ordenado con reemplazo

El tipo de muestreo ordenado con reemplazo es el más sencillo de describir. Aquí las muestras de tamaño $k \geq 1$ están ordenadas y con reemplazo. Note que k puede ser incluso más grande que el tamaño n de la población. Ante el reemplazo, un individuo de la población puede ser seleccionado hasta k veces. Este procedimiento se puede ver como el acomodo de k bolas en n cajas. La muestra de bolas selecciona a la población de cajas; como se aprecia en la siguiente figura.



La primera bola cabe en cualquiera de las n cajas, lo mismo que la segunda bola y el resto de ellas. Ante el reemplazo, una caja puede ser seleccionada hasta k veces. La selección de las cajas se representa en k casilleros, uno para cada bola. En cada casillero se escribe un número de caja; de n posibles, como se describe en la siguiente figura.



Con este esquema, el total de arreglos (muestras) posibles es

$$\#\Omega = n^k; \quad \text{con} \quad k, n \geq 1.$$

Cabe remarcar que el muestreo ordenado toma en cuenta la posición en que fueron seleccionados los individuos de la muestra. Además, el reemplazo hace factible un tamaño de muestra mayor que el de la población. Para este diseño muestral, no confundir casillero con caja, pues el arreglo de k casilleros representa a las k bolas.

Ejemplo 1.4.3. Se muestran ejemplos de muestreo ordenado con reemplazo.

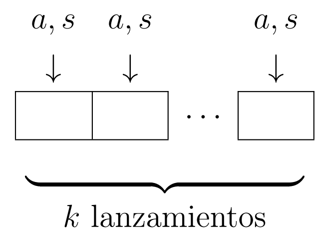
1. El Ejemplo 1.4.2.2, del lanzamiento de un dado cinco veces, describe un muestreo ordenado con reemplazo. En este caso, hay $k = 5$ lanzamientos y $n = 6$ caras. Por lo que

$$\#\Omega = n^k = 6^5 = 7776. \tag{1.25}$$

2. Otra aplicación del esquema en estudio es en las placas de automóvil de México, descrito en el Ejemplo 1.4.2.5. Las placas tienen $k = 9$ campos, mientras que la población se compone de $n = 38$ caracteres. El total de placas (muestras ordenadas sin reemplazo) posibles es

$$\#\Omega = n^k = 38^9 = 1.65 \times 10^{14}.$$

3. Cuando se lanza una moneda $k \geq 1$ veces, se tiene $n = 2$; como se describe en la siguiente figura.



El conjunto de eventos simples (muestras ordenadas con reemplazo) del experimento es

$$\Omega = \{(x_1, \ldots, x_k) : x_i = a, s\} = \{(x_1, \ldots, x_k) : x_i = 0, 1\}.$$

Su cardinalidad resulta

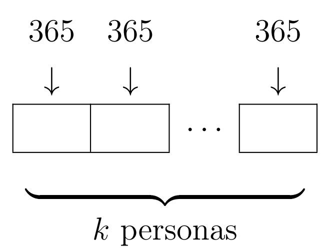
$$\#\Omega = n^k = 2^k.$$

En particular, si $k = 1$ entonces, hay $2^1 = 2$ resultados posibles del experimento. Ahora bien, el Ejemplo 1.2.4 describe el caso $k = 3$. Su espacio muestral es

$$\Omega = \{aaa, aas, asa, ass, saa, sas, ssa, sss\}.$$

Bajo el esquema de muestreo ordenado, las muestras aas y asa son diferentes. Note que el tamaño de muestra es ser mayor que el de la población.

4. Los cumpleaños de un grupo de k personas se puede representar con el esquema de bolas y cajas. En este sentido, hay $n = 365$ cajas y k bolas. En total, hay 365^k diferentes arreglos ordenados. Aquí el orden identifica a cada persona, como se describe en la siguiente figura.



5. De nuevo con el esquema de cajas y bolas, considere a un edificio de n pisos cuyo elevador parte con k pasajeros. Las diferentes maneras que suben las personas es n^k.

6. En una semana típica de la ciudad se registraron k accidentes de tránsito. Sus posibles frecuencias diarias pueden ser de $n^k = 7^k$ diferentes maneras. En el caso de $k = 32$ accidentes, un ejemplo de registro semanal de accidentes es $(2, 4, 4, 0, 5, 10, 7)$. Registro que difiere al de $(4, 2, 4, 0, 5, 10, 7)$. Aquí el día de la semana identifica al orden.

7. Con la secuencia AAB se distinguen las placas de taxis del Estado de Aguascalientes. La siguiente figura describe su esquema de muestreo ordenado y con reemplazo.

$$\begin{array}{ccccccccc} 10 & 10 & & 10 & 10 & & & & \\ \downarrow & \downarrow & & \downarrow & \downarrow & & & & \\ \hline \multicolumn{1}{|c|}{} & \multicolumn{1}{c|}{} & \multicolumn{1}{c|}{-} & \multicolumn{1}{c|}{} & \multicolumn{1}{c|}{} & \multicolumn{1}{c|}{-} & \multicolumn{1}{c|}{A} & \multicolumn{1}{c|}{A} & \multicolumn{1}{c|}{B} \\ \hline \end{array}$$

Bajo el esquema de cajas y bolas, hay $k = 4$ bolas y $n = 10$ cajas. Por el orden contemplado, la placa (arreglo o muestra) $43 - 22 - AAB$ es diferente a $34 - 22 - AAB$. El espacio muestral es

$$\Omega = \{(x_1, x_2, -, x_3, x_4, -, A, A, B) : x_i = 0, \ldots, 9,\ i = 1, 2, 3, 4\}.$$

Este se simplifica como un subconjunto de $\mathbf{R}^4$:

$$\Omega = \{(x_1, x_2, x_3, x_4) : x_i = 0, \ldots, 9,\ i = 1, 2, 3, 4\}.$$

El total de placas (muestras ordenadas con reemplazo) posibles para taxis es

$$\#\Omega = n^k = 10^4 = 10000.$$

1.4.2. Muestreo ordenado sin reemplazo: permutación

En esta sección se definen las funciones matemáticas factorial y permutación. Se define también el tipo de muestreo permutación.

El *factorial* es la función entera definida en los enteros no negativos como

$$n! = \begin{cases} 1 \times \cdots \times n & n \geq 1 \\ 1 & n = 0. \end{cases}$$

Más adelante notará la ventaja de definir $0! = 1$. Para $n \geq 1$, $n!$ es un producto de n enteros consecutivos. Este valor es el número de manera de ordenar a una población de n individuos, como se describe en la siguiente figura.

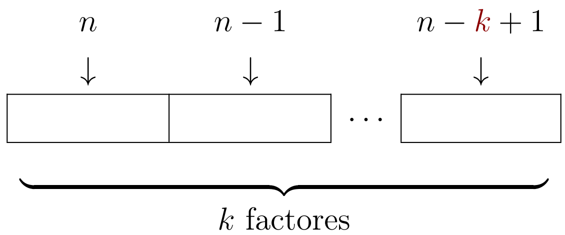
$$\underbrace{\begin{array}{ccccc} n & n-1 & & 2 & 1 \\ \downarrow & \downarrow & & \downarrow & \downarrow \\ \hline \multicolumn{1}{|c|}{} & \multicolumn{1}{c|}{} & \cdots & \multicolumn{1}{|c|}{} & \multicolumn{1}{c|}{} \\ \hline \end{array}}_{n \text{ factores}}$$

En el primer casillero se ubica al primer individuo seleccionado al azar de los n posibles. Fijado este, el segundo individuo se selecciona de los $n - 1$ restantes. Se continua así sucesivamente hasta el último individuo a ordenar.

El *coeficiente de permutación* es

$${}_nP_k = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1); \quad 0 \leq k \leq n.$$

Se lee literalmente o "n de k". La última igualdad aplica para $1 \leq k \leq n$, cuyo valor representa el total de *permutaciones* de tamaño k de una población de n individuos. El tipo de muestreo ordenado sin reemplazo se llama *permutación*. El orden de la muestra es relevante. Además, son diferentes los individuos, como lo describe el siguiente arreglo.



El primer individuo se escoge entre los n que integran la población. El segundo individuo se extrae de los $n-1$ restantes; que se acopla al primero. Proceda así sucesivamente hasta completar k individuos. Luego, el total de permutaciones es el coeficiente de permutación:

$$\#\Omega = {}_nP_k = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1); \quad \text{con} \quad 1 \leq k \leq n. \tag{1.26}$$

Como en la Sección 1.4.1, en este tipo de muestreo aplica el esquema de bolas y cajas, donde hay una población de n cajas, en la que se elige la muestra con k bolas. Sin embargo, aquí cada caja tiene a lo más una bola.

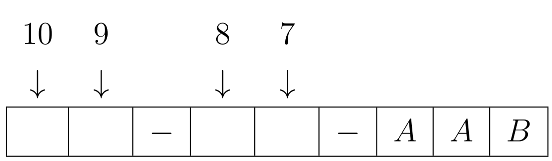
Ejemplo 1.4.4. Se describen algunas aplicaciones del muestreo por permutación.

1. El total de muestras de tamaño $k=1$ es n:
$$ {}_nP_1 = \frac{n!}{(n-1)!} = n. \tag{1.27}$$

2. Como arriba se mencionó, hay $n!$ permutaciones de una población de tamaño $n \geq 1$:
$$ {}_nP_n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!. \tag{1.28}$$

3. Por ejemplo, la población a, b, c, se permuta de $3! = 6$ diferentes maneras:
$$\Omega = \{abc, acb, bac, bca, cab, cba\}.$$
Si de esta población se extrae una permutación de tamaño $k=2$, entonces el total de muestras por permutación es
$$ {}_nP_k = {}_3P_2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = 6.$$
El correspondiente espacio muestral es
$$\Omega = \{ab, ac, ba, bc, ca, cb\}. \tag{1.29}$$

4. Retome el Ejemplo 1.4.3.7, de las placas de taxi. Suponga que un agente de tránsito observó a lo lejos un taxi que se pasó un crucero con la luz roja. Con la prisa, él sólo apreció que los dígitos de la placa no se repiten. La siguiente figura cuenta tal hecho.



El número de la placa del taxi infractor está entre los siguientes posibles:
$$ {}_{10}P_4 = \frac{n!}{(n-k)!} = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 10 \times 9 \times 8 \times 7 = 5\,040.$$

1.4.3. Muestreo no ordenado sin reemplazo: combinación

El *coeficiente binomial* se define como

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad 0 \leq k \leq n.$$

Se lee literalmente o "n en k". Es incorrecto mencionarlo como "n de k" o "n entre k"; pues son frases reservadas para otras funciones matemáticas. Una *combinación* es una muestra sin orden y sin reemplazo. Similar a la permutación, a falta de reemplazo, los individuos seleccionados son todos diferentes. Aunque, al omitir el orden, se reduce el total de muestras. Si $1 \leq k \leq n$, el total de muestras por combinación es el coeficiente binomial. Tanto el numerador como el denominador es un producto de k enteros consecutivos:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}.$$

Se describen algunas aplicaciones del muestreo por combinación. Note que

$$\binom{n}{0} = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1.$$

El total de muestras de tamaño $k = n$ es

$$\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!0!} = 1.$$

Contraste este valor con el total de permutaciones de la población: ${}_nP_n = n!$. Por otro lado, el total de muestras de tamaño $k = 1$ es

$$\binom{n}{1} = \frac{n!}{1!(n-1)!} = n.$$

Este número coincide con el total de muestras de tamaño $n-1$:

$$\binom{n}{n-1} = \frac{n!}{(n-1)!(n-(n-1))!} = \frac{n!}{(n-1)!\ 1!} = n.$$

De hecho, cada muestra de tamaño uno induce otra de tamaño $n-1$, y viceversa. En general, cada combinación de tamaño k induce otra combinación de tamaño $n-k$, y viceversa. Por lo que

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!}; \quad \text{para} \quad 0 \leq k \leq n. \tag{1.30}$$

Por ejemplo, si $n = 5$, sus coeficientes binomiales son

$$\binom{5}{0} = \binom{5}{5} = 1, \quad \binom{5}{1} = \binom{5}{4} = 5, \quad \binom{5}{2} = \binom{5}{3} = 10.$$

Hay 5 combinaciones de tamaño $k = 1$. Cada muestra de tamaño $k = 1$ induce otra de tamaño $k = 4$. Así mismo, hay 10 combinaciones de tamaño $k = 2$. Cada muestra de tamaño $k = 2$ induce otra de tamaño $k = 3$. Por otro lado, para $n = 6$, se tiene

$$\binom{6}{0} = \binom{6}{6} = 1, \quad \binom{6}{1} = \binom{6}{5} = 6, \quad \binom{6}{2} = \binom{6}{4} = 15, \quad \binom{6}{3} = 20.$$

$n=0$									1									
$n=1$								1		1								
$n=2$							1		2		1							
$n=3$						1		3		3		1						
$n=4$					1		4		6		4		1					
$n=5$				1		5		10		10		5		1				
$n=6$			1		6		15		20		15		6		1			
$n=7$		1		7		21		35		35		21		7		1		
$n=8$	1		8		28		56		70		56		28		8		1	

Tabla 1.11. Triángulo de Pascal, para $n = 0, \ldots, 8$ y $k = 0, \ldots, n$.

El *triángulo de Pascal* es la secuencia de coeficientes binomiales

$$\begin{aligned}
\binom{n}{0} &= \binom{n}{n} = 1, \\
\binom{n}{1} &= \binom{n}{n-1} = n, \\
\binom{n}{2} &= \binom{n}{n-2} = \frac{n(n-1)}{2}, \\
\binom{n}{3} &= \binom{n}{n-3} = \frac{n(n-1)(n-2)}{3!}, \\
&\vdots \\
\binom{n}{k} &= \binom{n}{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}; \quad \text{para} \quad 0 \leq k \leq n. \\
&\vdots
\end{aligned}$$

La Tabla 1.11 muestra los primeros elementos del triángulo de Pascal. Por (1.30), este es un arreglo simétrico respecto de k y $n-k$. Note también que el coeficiente binomial incrementa su valor conforme el entero k se acerca al centro $n/2$.

Note que cada combinación induce $k!$ permutaciones. Por lo tanto, la relación entre el total de combinaciones y permutaciones es

$$ {}_nC_k k! = {}_nP_k; \quad 0 \leq k \leq n.$$

Por (1.26), el total de combinaciones es el coeficiente binomial:

$$ {}_nC_k = \binom{n}{k} = \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!}; \quad 0 \leq k \leq n.$$

Ejemplo 1.4.5. Se muestran algunas aplicaciones del muestreo por combinación.

combinación	permutación					
abc	abc	acb	bac	bca	cab	cba
abd	abd	adb	bad	bda	dab	dba
abe	abe	aeb	bae	bea	eab	eba
acd	acd	adc	cad	cda	dac	dca
ace	ace	aec	cae	cea	eac	eca
ade	ade	aed	dae	dea	ead	eda
bcd	bcd	bdc	cbd	cdb	dbc	dcb
bce	bce	bec	cbe	ceb	ebc	ecb
bde	bde	bed	dbe	deb	ebd	edb
cde	cde	ced	dce	dec	ecd	edc

Tabla 1.12. Espacio muestral por combinación o por permutación, de una población de tamaño $n = 5$ y muestras de tamaño $k = 3$; Ejemplo 1.4.5.2.

1. Considere una muestra no ordenada sin reemplazo de tamaño $k = 2$, de la población a, b, c. Hay $\binom{3}{2} = 3$ diferentes combinaciones (muestras no ordenadas sin reemplazo):

$$\Omega = \{ab, ac, bc\}.$$

Si cada muestra se permuta (ordena), el espacio muestral de las permutaciones es (1.29). Por lo que

$${}_3P_2 = \binom{3}{2} \cdot 2! = \frac{3!}{2! \cdot 1!} \cdot 2 = 6.$$

Note que el espacio muestral Ω cambia si lo hace el tipo de muestreo. En este sentido, hay que tener precaución del esquema de muestreo aplicado.

2. Para la población a, b, c, d, e se obtendrán todas las combinaciones y permutaciones de tamaño $k = 3$. El conjunto de combinaciones posibles es

$$\Omega = \{abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde\}.$$

Así que

$$\#\Omega = \binom{n}{k} = \binom{5}{3} = \frac{5!}{3! \cdot 2!} = \frac{5 \times 4 \times 3!}{3! \times 2} = 10.$$

Cada combinación induce $k! = 3! = 6$ permutaciones; como se aprecia en la Tabla 1.12. El número de permutaciones es

$${}_nP_k = {}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60.$$

3. En la *Lotería Melate* se selecciona una combinación de tamaño $k = 6$ de una tómbola $n = 56$ bolas numeradas. El total de diferentes combinaciones es

$$\#\Omega = \binom{56}{6} = 32\,468\,436.$$

En consecuencia, el primer premio se gana con probabilidad

$$P(ganar) = \frac{1}{\#\Omega} = \frac{1}{\binom{56}{6}} = 5.55 \times 10^{-8}.$$

Compare este valor con la probabilidad de ganar en la Lotería Nacional: 2×10^{-5}; del Ejemplo 1.4.2.4. Ambos eventos son extremadamente poco probables de ocurrir. Sin embargo, el premio mayor de la Lotería Nacional es 649.37 más probable que el del Melate.

4. Considere una baraja de $n = 52$ cartas; véase el Ejemplo 1.4.2.6. Una *mano* es una combinación de $k = 5$ cartas. Cada mano es un evento simple del espacio muestral Ω. El total de manos es

$$\#\Omega = \binom{n}{k} = \binom{52}{5} = 2\,598\,960.$$

1.4.4. Muestreo no ordenado con reemplazo

La explicación de este esquema de muestreo es un poco más elaborada. El muestreo ordenado con reemplazo se describe como un arreglo de n casilleros que representan a los individuos de la población. Estos se seleccionan mediante k asteriscos. Cada individuo puede tener ninguno, uno, varios, hasta k asteriscos incluso. La siguiente figura describe este esquema de muestreo. El subíndice indica el número de individuo (casillero).

$$\overbrace{\boxed{\;**_1\;}\boxed{\quad\;_2\;}\boxed{\;***_3\;}\;\cdots\;\boxed{\quad *_n\;}}^{k \text{ asteriscos para } n \text{ casilleros}}$$

Cada muestra posible genera una secuencia de k asteriscos y $n-1$ barras interiores, que son las que separan a los casilleros. Las barras primera y última no participan, pues no se moverán. Ahora se contabilizará el total de secuencias de este tipo. Al considerar el orden, los $n+k-1$ objetos generan $(n+k-1)!$ secuencias de asteriscos y barras. Sin embargo, las secuencias de asteriscos (sin barras) no requieren estar ordenadas. Hay que omitir dicho orden, pues cada una de las $k!$ secuencias ordenadas de asteriscos son iguales. Lo mismo ocurre con las secuencias ordenadas de las barras (sin asteriscos). Su propio orden las replica $(n-1)!$ veces. Por lo que el total de arreglos no ordenados con reemplazo es

$$\frac{(n+k-1)!}{k!(n-1)!} = \binom{n+k-1}{n-1}.$$

Ejemplo 1.4.6. Se muestran ejemplos de muestreo ordenado con reemplazo.

1. Si de la población a, b, c, se toma una muestra de $k = 2$ individuos entonces, las muestras posibles se describen en la siguiente figura.

población			muestra
a	b	c	
$**$			aa
$*$	$*$		ab
$*$		$*$	ac
	$**$		bb
	$*$	$*$	bc
		$**$	cc

   La posición de los asteriscos y las barras interiores refleja a los individuos seleccionados. Note que el orden de los dos asteriscos no importa, así como el de las dos barras interiores. Esto confirma el número de muestras no ordenadas con reemplazo:

$$\binom{n+k-1}{n-1} = \frac{4!}{2! \cdot 2!} = 6.$$

2. Por otra parte, si $k = 3$ y $n = 5$ entonces, el número de muestras no ordenadas con reemplazo es

$$\binom{n+k-1}{n-1} = \binom{7}{3} = 35.$$

   Si denota como a, b, c, d, e a los miembros de esta población, el espacio muestral es

$$\begin{array}{ccccccc} aaa, & abb, & acd, & bbb, & bcd, & ccc, & cee, \\ aab, & abc, & ace, & bbc, & bce, & ccd, & ddd, \\ aac, & abd, & add, & bbd, & bdd, & cce, & dde, \\ aad, & abe, & ade, & bbe, & bde, & cdd, & dee, \\ aae, & acc, & aee, & bcc, & bee, & cde, & eee. \end{array}$$

3. El *cubilete* es una vaso con 5 dados cuyas caras son

7 ó 9 (negro)	8 ó 10 (rojo)	J (jota, sota, jack)
Q (reina, queen)	K (rey, king)	A (as, ace)

   Cada lanzamiento del cubilete es una realización del muestreo no ordenado con reemplazo. La población se representa por las $n = 6$ caras mientras que $k = 5$ es el total de dados. El total de jugadas diferentes del cubilete es

$$\#\Omega = \binom{n+k-1}{n-1} = \binom{6+5-1}{6-1} = \binom{10}{5} = 252.$$

	con reemplazo ($k, n \geq 1$)	sin reemplazo ($1 \leq k \leq n$)
ordenado	n^k	permutación: $_nP_k = (n)_k = \dfrac{n!}{(n-k)!}$
no ordenado	$\dbinom{n+k-1}{n-1}$	combinación: $_nC_k = \dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Tabla 1.13. Cardinalidad del espacio muestral según el tipo de muestreo.

Compare este resultado con la versión muestral ordenada, donde habría 7776 muestras ordenadas con reemplazo; véanse los ejemplos 1.4.2.2 y 1.4.3.1. Algunas anotaciones de los tipos de muestreo son las siguientes.

La Tabla 1.13 resume las fórmulas de conteo para los cuatro tipos básicos de muestreo, junto con las restricciones del tamaño de muestra k. Se comparten algunas anotaciones de los mismos.

1. El esquema de muestreo básico más usado es el muestreo por combinación. Esto se debe a que regularmente el orden de la muestra no es relevante. Además, por ser sin reemplazo, ofrece más información de la población a través de la muestra.

2. Para poblaciones muy grandes y tamaños de muestra relativamente pequeños, los tipos de muestreo con o sin reemplazo son en la práctica iguales. Esto se debe a la poca probabilidad que se repitan los individuos de la muestra.

3. Los arreglos vectoriales o casilleros aplican para los tipos de muestreo ordenados, ya sea con o sin reemplazo.

Con los cuatro tipos de muestreo básicos, se desarrollan otros esquemas más sofisticados; apoyándose en el teorema fundamental de conteo 1.4.1. Recuerde los casos descritos en los ejemplos 1.4.2.5-.6.

1.4.5. Muestreo hipergeométrico

En esta sección se describe el muestreo *hipergeométrico*, que es una generalización del muestreo por combinación.

Asuma que la población objetivo está separada, segregada o dividida por dos subpoblaciones o *estratos*, de tamaños respectivos n_1, $n_2 \geq 1$, con tamaño de la población global $n = n_1 + n_2$. Una muestra hipergeométrica establece submuestras en ambos estratos de la población. Para el primer estrato se selecciona una muestra por combinación de tamaño $0 \leq k_1 \leq n_1$. Análogamente, del segundo estrato seleccione una muestra por combinación de tamaño $0 \leq k_2 \leq n_2$. El tamaño de muestra global es $k = k_1 + k_2$, con $1 \leq k \leq n$. Por el primer estrato hay $\binom{n_1}{k_1}$ combinaciones mientras que $\binom{n_2}{k_2}$ es el total de combinaciones del segundo estrato. Cada combinación del primer estrato se acopla con su contraparte del segundo estrato. Del teorema fundamental de conteo 1.4.1, hay

$$\#\Omega = \binom{n_1}{k_1}\binom{n_2}{k_2}$$

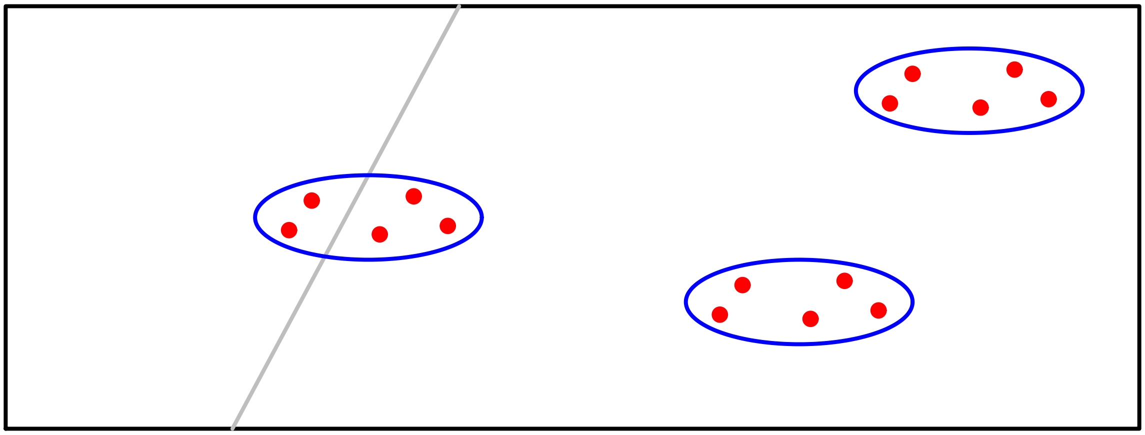


Figura 1.5. Población finita de dos estratos, junto con tres ejemplos de combinaciones de tamaño $k = 5$.

diferentes muestras hipergeométricas, bajo las restricciones

$$n = n_1 + n_2, \quad k = k_1 + k_2, \quad 1 \leq k \leq n, \quad 0 \leq k_1 \leq n_1, \quad 0 \leq k_2 \leq n_2.$$

El esquema de muestreo hipergeométrico permite al encuestador un control sobre el tamaño de las submuestras de cada estrato. Al asignar los tamaños de submuestras k_1 y k_2, se establece información muestral de cada uno de los estratos de la población. En caso contrario, es posible que la muestra por combinación tenga pocos individuos o ninguno de uno de los dos estratos, especialmente del de menor tamaño. En este sentido, el esquema de muestreo hipergeométrico mejora al de combinación. El diseño de muestreo hipergeométrico mejora la calidad de la información estadística sobre la o las variables de interés de la población. En la Figura 1.5 se aprecia una población finita de dos estratos, de donde se toman tres muestras por combinación de tamaño $k = 5$. Las muestras pueden incluir o no a individuos de ambos estratos. Recuerde que, con la población global de n individuos (sin estratos), hay $\binom{n}{k}$ combinaciones de tamaño k.

El muestreo hipergeométrico se extiende a una población dividida en $m \geq 1$ estratos, de tamaños $n_1, \ldots, n_m$. En este caso, el total de muestras hipergeométricas es

$$\#\Omega = \binom{n_1}{k_1} \cdots \binom{n_m}{k_m},$$

donde

$$\begin{aligned}
&n = n_1 + \cdots + n_m,\\
&k = k_1 + \cdots + k_m,\\
&1 \leq k \leq n,\\
&0 \leq k_1 \leq n_1, \ldots 0 \leq k_m \leq n_m
\end{aligned}$$

Ahora se describirán ejemplos del muestreo hipergeométrico.

Ejemplo 1.4.7. Considere una población de $n = 29$ personas, con $n_1 = 3$ mujeres y $n_2 = 26$ hombres. Se forma un comité de $k = 3$ personas seleccionadas al azar. Por la disparidad en el tamaño de los estratos, el muestreo por combinación puede dejar al comité sin representación de mujeres. El espacio muestral Ω es el conjunto de todas las combinaciones (muestras) posibles, con cardinalidad

$$\#\Omega = \binom{n}{k} = \binom{29}{3} = 3\,654.$$

El evento de interés es

$$\begin{aligned} B &: \text{las mujeres están representadas en el comité} \\ &= \text{al menos hay una mujer en el comité.} \end{aligned}$$

Para calcular su probabilidad, defina los siguientes eventos excluyentes

$$\begin{aligned} B_0 &: \text{comité de sólo hombres,} \\ B_1 &: \text{comité de una mujer y dos hombres,} \\ B_2 &: \text{comité de dos mujer y un hombre,} \\ B_3 &: \text{comité de sólo mujeres.} \end{aligned}$$

El subíndice del evento identifica el número de mujeres que tendría el comité. Note que B_0, B_1, B_2 y B_3 forman una partición de Ω, pues dichos eventos son excluyentes y su unión es Ω:

$$B_0 \cap B_1 = B_0 \cap B_2 = B_0 \cap B_3 = B_1 \cap B_2 = B_1 \cap B_3 = B_2 \cap B_3 = \varnothing$$

y

$$\Omega = B_0 \cup B_1 \cup B_2 \cup B_3.$$

Respecto del evento de interés, se tiene

$$B = B_1 \cup B_2 \cup B_3.$$

Por el tercer axioma de Kolmogorov (1.11), se tiene

$$P(B) = P(B_1 \cup B_2 \cup B_3) = P(B_1) + P(B_2) + P(B_3).$$

Para el evento B_1, de 3 mujeres se requiere una. Esto se realiza de $\binom{3}{1} = 3$ diferentes maneras. Así mismo, se piden 2 varones de los 26 disponibles, que corresponde a $\binom{26}{2} = 325$ diferentes combinaciones. Cada combinación del primer estrato se acopla con cada combinación del segundo. Por lo que

$$P(B_1) = \frac{\#B_1}{\#\Omega} = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} = \frac{\binom{3}{1}\binom{26}{2}}{\binom{29}{3}} = \frac{975}{3\,654} = 0.2668 = 26.68\,\%.$$

De forma similar, para el evento B_2, en el primer estrato hay $\binom{3}{2} = 3$ combinaciones de tamaño 2 mientras que del segundo corresponden $\binom{26}{1} = 26$ muestras de tamaño uno. Así

$$P(B_2) = \frac{\#B_2}{\#\Omega} = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} = \frac{\binom{3}{2}\binom{26}{1}}{\binom{29}{3}} = \frac{78}{3\,654} = 2.13 \times 10^{-2} = 2.13\,\%.$$

De forma análoga, se obtiene

$$P(B_3) = \frac{\#B_3}{\#\Omega} = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} = \frac{\binom{3}{3}\binom{26}{0}}{\binom{29}{3}} = \frac{1}{3\,654} = 2.74 \times 10^{-4} = 0.3\,\%.$$

Note que B_3 es un evento simple, pues hay sólo una muestra hipergeométrica a su favor. De esta manera

$$\begin{aligned} P(B) &= P(B_1 \cup B_2 \cup B_3) \\ &= P(B_1) + P(B_2) + P(B_3) \\ &= 0.2668 + 2.13 \times 10^{-2} + 2.74 \times 10^{-4} \\ &= 0.2884 = 28.84\,\%. \end{aligned} \tag{1.31}$$

El evento de que las mujeres estén representadas en el comité tiene una probabilidad de ocurrir 0.2884; la cual resulta baja pero no muy chica. Su complemento es más factible de ocurrir:

$$P(B^c) = 1 - P(B) = 1 - 0.2884 = 0.7116 = 71.16\,\%.$$

Note que

$$B^c = (B_1 \cup B_2 \cup B_3)^c = B_0,$$
$$\#B_0 = \binom{n_1}{k_1}\binom{n_2}{k_2} = \binom{3}{0}\binom{26}{3}$$

y

$$P(B_0) = \frac{\#B_0}{\#\Omega} = \frac{\binom{3}{0}\binom{26}{3}}{\binom{29}{3}} = 0.7116 = 71.16\,\%.$$

Por lo tanto, hay otra vía para calcular la probabilidad del evento de interés:

$$P(B) = P(B_0^c) = 1 - P(B_0) = 1 - 0.7116 = 0.2884 = 28.84\,\%.$$

Por otro lado, se confirma la ley de probabilidad total (1.19), aplicada a la partición de Ω:

$$\begin{aligned} P(\Omega) &= P(B_0) + P(B_1) + P(B_2) + P(B_3) \\ &= 0.7116 + 0.2668 + 2.13 \times 10^{-2} + 2.74 \times 10^{-4} \\ &= 1. \end{aligned}$$

Una aplicación de la ley de probabilidad total es la fórmula

$$\binom{29}{3} = \binom{3}{0}\binom{26}{3} + \binom{3}{1}\binom{26}{2} + \binom{3}{2}\binom{26}{1} + \binom{3}{3}\binom{26}{0}.$$

El lado izquierdo representa el total de combinaciones de tamaño $k = 3$ de una población de $n = 29$ personas. Los términos del lado derecho consideran a todos los posibles muestreos hipergeométricos, donde los tamaños de muestra k_1 y k_2 son tal que $k = k_1 + k_2 = 3$. Por el Ejercicio 1.19, la fórmula general que relaciona las combinaciones con todos los posibles muestreos hipergeométricos de dos estratos es

$$\binom{n}{k} = \binom{n_1}{0}\binom{n_2}{k} + \binom{n_1}{1}\binom{n_2}{k-1} + \cdots + \binom{n_1}{k-1}\binom{n_2}{1} + \binom{n_1}{k}\binom{n_2}{0},$$

donde $n = n_1 + n_2$ y $0 \leq k \leq n_1, n_2$.

Ejemplo 1.4.8. En este ejemplo se analizarán las manos del juego de la baraja: par, dos pares, tercia, escalera, flor, full, póquer y flor con escalera. Salvo por la flor, la definición de estas manos se basa en primer término por el número de la carta y luego por su figura. En este sentido, y en el contexto del teorema fundamental de conteo, se recomienda primero contar las tareas asociadas al número, para luego hacer lo propio con las tareas de figura. El experimento consiste en seleccionar al azar una mano, que es una combinación de $k = 5$ cartas de las $n = 52$ posibles. El total de manos (combinaciones) en el juego de la baraja es

$$\#\Omega = \binom{n}{k} = \binom{52}{5} = 2\,598\,960.$$

1. ¿Cuál es la probabilidad de obtener un *póquer de ases*? La mano póquer de ases es:

   $$AAAAy, \quad \text{con} \quad y \neq A.$$

   La carta y, que difiere al as, puede ser cualquiera de las 48 restantes. Por lo que

   $$\#\text{póquer de ases} = 48$$

   y

   $$P(\text{póquer de ases}) = \frac{\#\text{póquer de ases}}{\#\Omega} = \frac{48}{\binom{52}{5}} = 1.84 \times 10^{-5}. \tag{1.32}$$

   Otra manera de contar la carta y es elegir primero su número y luego la figura. El número y se elige de los 12 posibles. Luego se elige una figura de las 4 posibles. La siguiente tabla describe el conteo.

   $$\begin{array}{cc} 12 & 4 \\ \downarrow & \downarrow \\ \boxed{y} & \boxed{\text{figura de } y} \end{array}$$

Lo que ratifica la cardinalidad del evento de interés:

$$\#\text{póquer de ases} = 12 \times 4 = 48.$$

Ahora se analizará el *póquer*, que es una mano con cuatro números iguales:

$$xxxxy; \quad \text{con} \quad x \neq y.$$

Primero contabilice la elección de los números x y y, para luego elegir sus figuras. La mano póquer involucra dos números diferentes de los 13 posibles. Aquí hay que cuestionarse si aplica el muestreo por permutación o combinación, según si es relevante o no el orden. El orden de elección de x y y es importante pues, al intercambiar su rol, son diferentes las manos $xxxxy$ y $yxxxx$. El número x se selecciona de 13 posibles mientras que para y es cualquiera de los 12 restantes:

$$_{13}P_2 = \frac{13!}{(13-2)!} = \frac{13!}{11!} = 13 \times 12.$$

Luego de fijar los números x, y, hay que contabilizar sus figuras. Para ello, aplique el muestreo hipergeométrico, donde la población de 52 cartas se separa en tres estratos: 4 cartas con número x, 4 cartas con número y, así como el resto de las 44 cartas. El total de muestras hipergeométricas del evento de interés es $\binom{4}{4}\binom{4}{1}\binom{44}{0}$. Por lo que

$$\#\text{póquer} = 13 \times 12 \times \binom{4}{4}\binom{4}{1}\binom{44}{0} = 624.$$

La siguiente tabla resume el calculo de la cardinalidad del evento de interés.

$$\begin{array}{cccc}
13 & \binom{4}{4} & 12 & \binom{4}{1} \\
\downarrow & \downarrow & \downarrow & \downarrow \\
\boxed{x} & \boxed{\text{figura de } x} & \boxed{y} & \boxed{\text{figura de } y}
\end{array}$$

Así

$$P(\text{póquer}) = \frac{\#\text{póquer}}{\#\Omega} = \frac{624}{\binom{52}{5}} = 0.0002 = 0.02\,\%.$$

Otra ruta para calcular esta probabilidad es multiplicar por 13 la probabilidad del póquer de ases (1.32). De hecho, el evento póquer es la unión de 13 eventos excluyentes y equiprobables: póquer de ases, dos,..., o reyes.

2. Un *full* es una mano con tres cartas del mismo número junto con dos de otro número (tercia con par):

$$xxxyy; \quad \text{con} \quad x \neq y.$$

Primero se analizará el ejemplo particular de full de reyes y ochos: $KKK88$. Para escoger las figuras de los números K y 8, separe a la población de 52 cartas en tres estratos: reyes, ochos y el resto de las cartas. De 4 reyes se toman 3, de 4 ochos escoger 2 y cero cartas del resto. Por lo que

$$P(KKK88) = \frac{\binom{4}{3}\binom{4}{2}\binom{44}{0}}{\binom{52}{5}} = \frac{24}{\binom{52}{5}} = 9.23 \times 10^{-6}.$$

El orden de los números importa, pues $KKK88 \neq 888KK$. Así mismo, para el full general, se tiene $xxxyy \neq yyyxx$. Como en el póquer, la mano full requiere dos números diferentes y ordenados, es decir, permutación. Los números x, y se obtienen por permutación, de 13 números elegir 2: ${}_{13}P_2 = 13 \times 12$. Al fijar los números x, y, ahora cuente sus figuras. La siguiente tabla describe el total de eventos simples a favor del full.

$$\begin{array}{cccc} 13 & \binom{4}{3} & 12 & \binom{4}{2} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{x} & \boxed{\text{figura de } x} & \boxed{y} & \boxed{\text{figura de } y} \end{array}$$

Se concluye

$$P(\text{full}) = \frac{13\binom{4}{3} \times 12\binom{4}{2}}{\binom{52}{5}} = \frac{3\,744}{\binom{52}{5}} = 0.0014 = 0.14\,\%.$$

Tanto las manos póquer como full son eventos poco probables. Estos ocurren 2 y 14 de cada diez mil manos; respectivamente. Sin embargo, el full es 6 veces más factible que el póquer. Esta comparación de eventos raros aplica en el contexto de medicina, por ejemplo, para comparar el riesgo de contagio de dos enfermedades graves.

3. Una *tercia* es una mano de la forma $xxxyz$, donde x, y y z son diferentes:

$$xxxyz; \quad x \neq y \neq z \neq x.$$

La tercia involucra tres números diferentes, de modo que descarta las manos full y póquer. Se analizará el caso particular de tercia de reinas:

$$QQQyz; \quad \text{con} \quad Q \neq y \neq z \neq Q.$$

Como $QQQyz = QQQzy$, entonces no importa el orden de los números y, z. Su elección se realiza de $\binom{12}{2}$ diferentes maneras. En el siguiente paso se determinan las figuras de Q, y, z. Por el muestreo hipergeométrico, esto se realiza de $\binom{4}{3}\binom{4}{1}\binom{4}{1}\binom{40}{0}$ maneras. La siguiente tabla describe el conteo.

$$\begin{array}{cccc} \binom{4}{3} & \binom{12}{2} & \binom{4}{1} & \binom{4}{1} \\ \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{\text{figura de } Q} & \boxed{y, z} & \boxed{\text{figura de } y} & \boxed{\text{figura de } z} \end{array}$$

Así

$$P(\text{tercia de reinas}) = \frac{\binom{12}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} = \frac{4224}{\binom{52}{5}} = 0.0016 = 0.16\,\%. \qquad (1.33)$$

Para la tercia general, son iguales las manos $xxxyz$ y $xxxzy$, pero diferentes a $yyyxz$ o $zzzyx$. Elija primero el número x, que se realiza de 13 diferentes maneras. Luego, de los 12 números restantes se eligen y y z, cuyo orden no importa; así que es una combinación: $\binom{12}{2}$. Por último, la elección de las figuras de x, y y z se obtiene de $\binom{4}{3}$, $\binom{4}{1}$ y $\binom{4}{1}$ diferentes maneras; respectivamente. La siguiente tabla describe el conteo.

$$\begin{array}{ccccc} 13 & \binom{4}{3} & \binom{12}{2} & \binom{4}{1} & \binom{4}{1} \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{x} & \boxed{\text{figura de } x} & \boxed{y, z} & \boxed{\text{figura de } y} & \boxed{\text{figura de } z} \\ \uparrow & & \uparrow & & \\ 11 & & \binom{13}{2} & & \end{array}$$

Esta tabla describe también un conteo alternativo de las tareas asociadas al número. Primero se escoge al par de números y y z, el cual se realiza de $\binom{13}{2}$ diferentes maneras. Luego, la selección del número x se realiza con los 11 restantes. De hecho, la equivalencia de ambos conteos se verifica por lo siguiente:

$$13 \times \binom{12}{2} = 11 \times \binom{13}{2}.$$

Se concluye

$$P(\text{tercia}) = \frac{13\binom{4}{3} \times \binom{12}{2}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} = \frac{54\,912}{\binom{52}{5}} = 0.0211 = 2.11\,\%.$$

4. La mano *dos pares* es el arreglo $xxyyz$; con x, y, z diferentes:

$$xxyyz; \quad x \neq y \neq z \neq x.$$

De nuevo, esta mano involucra tres números diferentes. El orden entre los números x, y no importa, pues son iguales las manos $xxyyz$ y $yyxxz$. El par de números se elige de los 13 disponibles: $\binom{13}{2}$. Luego, el número z se escoge de los 11 restantes, como se describe en la siguiente tabla.

$$
\begin{array}{ccccc}
\binom{13}{2} & \binom{4}{2} & \binom{4}{2} & 11 & \binom{4}{1} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\boxed{x, y} & \boxed{\text{figura de } x} & \boxed{\text{figura de } y} & \boxed{z} & \boxed{\text{figura de } z} \\
\uparrow & & & \uparrow & \\
\binom{12}{2} & & & 13 &
\end{array}
$$

Esta misma tabla describe el conteo alternativo, donde se elige primero al número z de 13 maneras diferentes, para luego elegir a x, y de $\binom{12}{2}$ diferentes maneras. De hecho

$$\binom{13}{2} \times 11 = \binom{12}{2} \times 13.$$

Se concluye

$$P(\text{dos pares}) = \frac{\binom{13}{2}\binom{4}{2}\binom{4}{2} \times 11 \binom{4}{1}}{\binom{52}{5}} = \frac{123\,552}{\binom{52}{5}} = 0.0475 = 4.75\,\%.$$

Cabe mencionar que, por definición, la mano dos pares excluye a póquer o full.

5. La mano con jugada más factible es el *par*. Este es el arreglo

$$xxyzu; \quad x \neq y \neq z \neq x \neq u \neq y, \quad z \neq u.$$

El par excluye a las manos dos pares, tercia, full y póquer. El número x se elige de 13 posibles. Luego, los números y, z, u se escogen por combinación: $\binom{12}{3}$. La cardinalidad del evento de interés se explica en la siguiente tabla.

$$
\begin{array}{cccccc}
13 & \binom{4}{2} & \binom{12}{3} & \binom{4}{1} & \binom{4}{1} & \binom{4}{1} \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\boxed{x} & \boxed{\text{figura de } x} & \boxed{y, z, u} & \boxed{\text{figura de } y} & \boxed{\text{figura de } z} & \boxed{\text{figura de } u}
\end{array}
$$

Por lo tanto

$$P(\text{par}) = \frac{13\binom{4}{2} \times \binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} = \frac{1\,098\,240}{\binom{52}{5}} = 0.4226 = 42.26\,\%.$$

6. Una *flor con escalera* es una secuencia de 5 números consecutivos con la misma figura, por ejemplo: $A\heartsuit 2\heartsuit 3\heartsuit 4\heartsuit 5\heartsuit$ o $9\heartsuit 10\heartsuit J\heartsuit Q\heartsuit K\heartsuit$. El conteo se realiza de forma similar a los eventos arriba descritos. En ese sentido, primero seleccione el número inicial x de la secuencia, para luego acoplarlo con su figura. El primer número se selecciona de $9 + 1$ posibles. Aquí se toma en cuenta que el as también tiene rol de número 14; por ejemplo $10\diamondsuit J\diamondsuit Q\diamondsuit K\diamondsuit A\diamondsuit$. El siguiente paso es acoplar la secuencia numérica con cada una de las 4 figuras; como se describe en la siguiente tabla.

$$\begin{array}{cc} 9+1 & 4 \\ \downarrow & \downarrow \\ \boxed{x} & \boxed{\text{figura de } x} \end{array}$$

Así que

$$P(\text{flor con escalera}) = \frac{(9+1)4}{\binom{52}{5}} = \frac{40}{\binom{52}{5}} = 1.54 \times 10^{-5}.$$

7. La *escalera* es una secuencia de 5 números consecutivos, que no sea flor con escalera:

$$x(x+1)(x+2)(x+3)(x+4).$$

Algunos ejemplos de escalera son las manos

$$A\heartsuit 2\heartsuit 3\spadesuit 4\clubsuit 5\diamondsuit, \quad 9\heartsuit 10\heartsuit J\clubsuit Q\spadesuit K\diamondsuit \quad \text{y} \quad 10\diamondsuit J\heartsuit Q\spadesuit K\clubsuit A\heartsuit.$$

El primer número de la secuencia se escoge de 10 diferentes maneras. Cada número de la secuencia dispone de cuatro figuras; como se indica en la siguiente tabla.

$$\begin{array}{cccccc} 10 & 4 & 4 & 4 & 4 & 4 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ \boxed{x} & \boxed{\text{fig. de } x} & \boxed{\text{fig. de } x+1} & \boxed{\text{fig. de } x+2} & \boxed{\text{fig. de } x+3} & \boxed{\text{fig. de } x+4} \end{array}$$

El conteo incluye a las 40 manos de flor con escalera, las cuales deben descontarse. El total de eventos simples a favor de la escalera es

$$\#\text{escalera} = 10 \times 4^5 - 40 = 10200.$$

Aquí el factor 4^5 corresponde a un muestreo ordenado con reemplazo de la elección de figura. Por lo que

$$P(\text{escalera}) = \frac{10\,200}{\binom{52}{5}} = 0.0039 = 0.39\,\%.$$

8. Una *flor* es una mano de igual figuras, que no sea flor con escalera. Por ejemplo, una flor de corazones es la mano $3\heartsuit 5\heartsuit 7\heartsuit 8\heartsuit 10\heartsuit$. Aquí la figura tiene el rol principal. Así que primero seleccione la figura de las cartas, para luego decidir sobre su número. Hay cuatro figuras. Imagine de momento que la figura seleccionada es el corazón $\heartsuit$. Ahora seleccione por combinación 5 números de 13, como se describe en la siguiente tabla.

$$\begin{array}{cc} 4 & \binom{13}{5} \\ \downarrow & \downarrow \\ \boxed{\text{figura}} & \boxed{x, y, z, u, v} \end{array}$$

Como en el evento escalera, se quitan las 40 manos de flor con escalera. Por lo que

$$P(\text{flor}) = \frac{4\binom{13}{5} - 40}{\binom{52}{5}} = \frac{5\,108}{\binom{52}{5}} = 0.002 = 0.20\,\%.$$

9. La Tabla 1.14 resumen las cardinalidades y probabilidades de las manos descritas en este ejemplo, las cuales son excluyentes. Al sumar sus probabilidades, se obtiene la probabilidad de una jugada:

$$\begin{aligned} P(\text{jugada}) &= P(\text{par}) + P(\text{dos pares}) + P(\text{tercia}) + P(\text{escalera}) \\ &\quad + P(\text{flor}) + P(\text{full}) + P(\text{póquer}) + P(\text{flor con escalera}) \\ &= \frac{10981240 + 123552 + 54912 + 10200 + 5108 + 3744 + 624 + 40}{2598960} \\ &= 0.4988. \end{aligned}$$

Su complemento es un poco más factible:

$$P(\text{sin jugada}) = 1 - P(\text{jugada}) = 1 - 0.4987 = 0.5012.$$

10. Una alternativa de verificación de la probabilidad de la mano sin jugada es la siguiente. Denote como B_4 al evento mano sin jugada. Este evento está contenido en el evento B de todas las manos con números diferentes:

$$xyzuv; \quad \text{todos diferentes.}$$

Note que B es la unión de cuatro eventos excluyentes:

$$B = B_1 \cup B_2 \cup A_8 \cup B_4,$$

donde los eventos B_1, B_2 y B_3 son las manos flor con escalera, escalera y flor; respectivamente. Entonces

$$\#B = \#(B_1 \cup B_2 \cup B_3 \cup B_4) = \#B_1 + \#B_2 + \#B_3 + \#B_4,$$

donde

$$\#B = \binom{13}{5}\binom{4}{1}\binom{4}{1}\binom{4}{1}\binom{4}{1}\binom{4}{1} = 1317888$$

y

$$\#B_1 + \#A_7 + \#B_3 = 40 + 10200 + 5108 = 15348.$$

Así

$$\#B_1 = 1317888 - 15348 = 1\,302\,540.$$

mano (evento)	cardinalidad	probabilidad
flor con escalera	$(9+1)\times 4 = 40$	< 0.0001
póquer	$13 \times 48 = 624$	0.0002
full	$13\binom{4}{3} \times 12\binom{4}{2} = 3744$	0.0014
flor	$4\binom{13}{5} - 40 = 5108$	0.0020
escalera	$10\binom{4}{1}^5 - 40 = 10200$	0.0039
tercia	$13\binom{4}{3} \times \binom{12}{2}\binom{4}{1}^2 = 54912$	0.0211
dos pares	$\binom{13}{2}\binom{4}{2}^2 \times 11\binom{4}{1} = 123552$	0.0475
par	$13\binom{4}{2} \times \binom{12}{3}4^3 = 1098240$	0.4226
con jugada (anteriores)	1296420	0.4987
sin jugada (complemento)	$\binom{52}{5} - 1296420 = 1302540$	0.5012
todas las manos	$\#\Omega = \binom{52}{5} = 2598960$	1

Tabla 1.14. Algunas manos del juego de cartas de póquer, junto con su cardinalidad y probabilidad; Ejemplo 1.4.8.

Ejemplo 1.4.9. En un club de bridge de 6 parejas se forma un comité de 4 personas seleccionadas al azar. Encuentre la probabilidad de los siguientes eventos.
1. El comité se compone de dos hombres y dos mujeres
2. Los miembros del comité son del mismo sexo
3. En el comité hay al menos dos hombres
4. El comité es de dos hombres y dos mujeres, sin parejas originales
5. El comité no contiene parejas originales.

Solución. El espacio muestral Ω se compone de la colección de todos los comités posibles (muestras por combinación):

$$\#\Omega = \binom{n}{k} = \binom{12}{4} = 495.$$

Para resolver este problema defina los siguientes eventos excluyentes:

$$\begin{aligned}
A_0 &: \text{hay sólo mujeres en el comité,}\\
A_1 &: \text{hay exactamente un hombre en el comité,}\\
A_2 &: \text{hay exactamente dos hombres en el comité,}\\
A_3 &: \text{hay exactamente tres hombres en el comité,}\\
A_4 &: \text{hay sólo hombres en el comité.}
\end{aligned}$$

El subíndice del evento identifica al número de hombres que tendría el comité. Como en el Ejemplo 1.4.7, la construcción de eventos excluyentes facilita la aplicación de la ley de probabilidad total. Note además que los eventos A_0, A_1, A_2, A_3 y A_4 forman una partición del espacio muestral Ω:

$$\Omega = A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4, \quad \text{con}$$
$$A_i \neq A_j; \quad \text{para} \quad i \neq j, \quad i, j = 0, 1, 2, 3, 4.$$

1. El evento de interés es A_2. Al invocar al muestreo hipergeométrico, estratifique la población en 6 hombres y 6 mujeres, de donde se escogen 2 hombres y 2 mujeres; respectivamente. Así

$$P(A_2) = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n}{k}} = \frac{\binom{6}{2}\binom{6}{2}}{\binom{12}{4}} = \frac{225}{495} = 0.4545 = 45.45\,\%.$$

2. Aquí el evento de interés es la unión de eventos excluyentes $B = A_0 \cup A_4$. Que el comité sea de sólo mujeres se realiza de $\binom{6}{0}\binom{6}{4}$ diferentes maneras. Un número igual de muestras aporta un comité de sólo hombres: $\binom{6}{4}\binom{6}{0}$. Por lo que

$$\begin{aligned} P(B) &= P(A_0 \cup A_4) = P(A_0) + P(A_4) \\ &= \frac{\binom{6}{0}\binom{6}{4} + \binom{6}{4}\binom{6}{0}}{\binom{12}{4}} = \frac{15 + 15}{495} \\ &= 0.0606 = 6.06\,\%. \end{aligned}$$

3. Que haya al menos dos hombres se representa como la unión de tres eventos excluyentes: $C = A_2 \cup A_3 \cup A_4$. Por lo que

$$\begin{aligned} P(C) &= P(A_2 \cup A_3 \cup A_4) = P(A_2) + P(A_3) + P(A_4) \\ &= \frac{\binom{6}{2}\binom{6}{2} + \binom{6}{3}\binom{6}{1} + \binom{6}{4}\binom{6}{0}}{\binom{12}{4}} = \frac{225 + 120 + 15}{495} \\ &= 0.7272 = 72.72\,\%. \end{aligned}$$

Para otra solución de este problema, cuantifique la probabilidad del complemento:

$$C^c = (\text{al menos hay dos hombres})^c = \text{hay menos de dos hombres} = A_0 \cup A_1$$

Así que

$$\begin{aligned} P(C^c) = P(A_0 \cup A_1) &= \frac{\binom{6}{0}\binom{6}{4} + \binom{6}{1}\binom{6}{3}}{\binom{12}{4}} \\ &= \frac{15 + 120}{495} = 0.2727. \end{aligned}$$

y

$$P(C) = 1 - P(C^c) = 1 - 0.2727 = 0.7273.$$

4. En el evento de interés D, se requiere un comité de 2 hombres y 2 mujeres, pero que sean de parejas diferentes. Primero seleccione a los hombres para luego hacer lo propio con las mujeres. Tome una combinación de 2 hombres de los 6 disponibles: $\binom{6}{2}$. Sus respectivas parejas se descartan para la elección de las mujeres requeridas. Esto se realiza de $\binom{4}{2}$ diferentes maneras. Por lo que

$$P(D) = \frac{\#D}{\#\Omega} = \frac{\binom{6}{2}\binom{4}{2}}{\binom{12}{4}} = \frac{90}{495} = 0.1818 = 18.18\,\%.$$

5. El evento E, un comité sin parejas originales, se separa como sigue

$$\begin{aligned} E &= E \cap \Omega = E \cap (A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4) \\ &= (E \cap A_0) \cup (E \cap A_1) \cup (E \cap A_2) \cup (E \cap A_3) \cup (E \cap A_4), \end{aligned}$$

donde

$$\begin{aligned} &E \cap A_0 = A_0, \\ &E \cap A_1 : \text{hay exactamente un hombre y sin su pareja}, \\ &E \cap A_2 : \text{hay exactamente dos hombres y sin sus parejas}, \\ &E \cap A_3 : \text{hay exactamente tres hombres y sin sus parejas}, \\ &E \cap A_4 = A_4. \end{aligned}$$

Por la ley de probabilidad total (1.19), se tiene

$$\begin{aligned} P(E) &= P((E \cap A_0) \cup (E \cap A_1) \cup (E \cap A_2) \cup (E \cap A_3) \cup (E \cap A_4)) \\ &= P(A_0) + P(E \cap A_1) + P(E \cap A_2) + P(E \cap A_3) + P(A_4) \\ &= \frac{\binom{6}{0}\binom{6}{4} + \binom{6}{1}\binom{5}{3} + \binom{6}{2}\binom{4}{2} + \binom{6}{3}\binom{3}{1} + \binom{6}{4}\binom{6}{0}}{\binom{12}{4}} \\ &= 0.4849 = 48.49\,\%. \end{aligned}$$

Este problema se resolvió con la partición del evento de interés E, para luego aplicar la ley de probabilidad total. Ahora se abordará otro enfoque. Un comité de 4 personas sin sus parejas originales involucra sólo a un miembro de dichas 4 parejas, ya sea hombre o mujer, pero no a ambos. Así que primero seleccione a 4 parejas de las 6 disponibles: $\binom{6}{4}$. El siguiente paso consiste en elegir a un hombre o mujer de cada pareja seleccionada, lo que resulta de $2 \times 2 \times 2 \times 2$ diferentes maneras. Por lo que

$$P(E) = \frac{\binom{6}{4}2^4}{\binom{12}{4}} = 0.4849.$$

Por último, compare las probabilidades de los eventos excluyentes A_0, A_1, A_2, A_3 y A_4. El evento más probable es A_2, con $P(A_2) = 0.4545$. De hecho, como los estratos de hombres y mujeres son del mismo tamaño, este es el evento esperado. Le siguen los eventos A_1 y A_3, con igual probabilidad: $P(A_1) = P(A_3) = 0.2424$. Finalmente, en pocas ocasiones ocurrirán los eventos A_0 y A_4, pues $P(A_0) = P(A_4) = 0.03$.

1.5. Independencia de eventos

En esta sección se describirá la independencia de eventos.
Dos eventos A y B son *independientes* si

$$P(A \cap B) = P(A)P(B). \tag{1.34}$$

Se denota como $A \perp B$. En caso contrario, se dice que A y B son eventos *dependientes*.

Ejemplo 1.5.1 (continuación de Ejemplo 1.2.5)**.** Considere un dado legal que se lanza dos veces y registra el número de su cara superior; véase el espacio muestral Ω en las tablas 1.5 y 1.15. Sean los eventos

$$\begin{aligned}
A_2 &: \text{cara superior del primer lanzamiento es } 2,\\
A_5 &: \text{cara superior del primer lanzamiento es } 5,\\
B_5 &: \text{cara superior del segundo lanzamiento es } 5,\\
D &: \text{cara superior del primer lanzamiento es } 2 \text{ ó } 5 = A_2 \cup A_5.
\end{aligned}$$

Sus respectivas probabilidades son

$$\begin{aligned}
P(A_2) = P(A_5) = P(B_5) &= \frac{6}{36} = \frac{1}{6},\\
P(D) &= \frac{12}{36} = \frac{1}{3}.
\end{aligned}$$

En la Tabla 1.15 se aprecian los eventos de interés A_2, A_5 y B_5. Como el resultado del primer lanzamiento no influye sobre el segundo, y viceversa, son independientes los eventos A_2 y B_5. El evento A_2 no favorece ni perjudica las posibilidades del evento B_5. Viceversa, la ocurrencia del evento B_5 no facilita ni inhibe las posibilidades del evento A_2. De hecho

$$P(A_2 \cap B_5) = P(\{(2,5)\}) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(A_2)P(B_5).$$

De la misma manera, se verifica la independencia entre los eventos A_5 y B_5:

$$P(A_5 \cap B_5) = P(\{(5,5)\}) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(A_5)P(B_5).$$

Así mismo, son independientes los eventos D y B_5:

$$P(D \cap B_5) = P(\{(2,5),(5,5)\}) = \frac{2}{36} = \frac{2}{6} \cdot \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{6} = P(D)P(B_5).$$

Ω	1	2	3	4	5	6	
1	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$\cancel{(1,5)}$	$(1,6)$	
2	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$\cancel{(2,5)}$	$(2,6)$	A_2
3	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$\cancel{(3,5)}$	$(3,6)$	
4	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$\cancel{(4,5)}$	$(4,6)$	
5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$\cancel{(5,5)}$	$(5,6)$	A_5
6	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$\cancel{(6,5)}$	$(6,6)$	B_5

Tabla 1.15. Eventos A_2 amarillo, A_5 rosa y B_5 raya roja; Ejemplo 1.5.1.

En general, hay independencia entre cualquier evento asociado sólo al primer lanzamiento y otro sólo del segundo. En contraste, son dependientes los eventos A_2 y A_5. De hecho, como son excluyentes, son muy dependientes. Si ocurrió A_2, se descarta el evento A_5, y viceversa. Si no ocurrió A_2, es decir se tiene el evento A_2^c, entonces favorece la ocurrencia del evento A_5. Formalmente, la dependencia se verifica como sigue

$$P(A_2 \cap A_5) = P(\varnothing) = 0 \neq \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = P(A_2)P(A_5)$$

y

$$P(A_2^c \cap A_5) = P(A_5) = \frac{1}{6} \neq \frac{5}{36} = \frac{5}{6} \times \frac{1}{6} = P(A_2^c)P(A_5).$$

La primera igualdad de la última expresión se debe a que $A_5 \subset A_2^c$. De la misma manera, se verifica la dependencia entre los eventos A_2 y D:

$$P(A_2 \cap D) = P(A_2) = \frac{1}{6} > \frac{1}{6} \cdot \frac{1}{3} = P(A_2)P(D).$$

La primera igualdad de esta expresión se debe a que $A_2 \subset D$.

El concepto de independencia se extiende a sucesiones de dos o más eventos. Se dice que los eventos A, B y C son *independientes* si son independientes por pares y

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{1.35}$$

Junto con (1.35), se deben cumplir las igualdades

$$P(A \cap B) = P(A)P(B), \tag{1.36}$$
$$P(A \cap C) = P(A)P(C), \tag{1.37}$$
$$P(B \cap C) = P(B)P(C). \tag{1.38}$$

Cabe mencionar que hay ejemplos de eventos A, B y C, que satisfacen (1.35), pero no (1.36)-(1.38). Con lo que dichos eventos no serán independientes; véase el Ejercicio 1.22.

En general, son *independientes* los eventos $A_1, A_2, \ldots, A_n$, si se satisface

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n) \tag{1.39}$$

y son independientes toda subcolección de $n-1$ eventos. Una definición equivalente del concepto de independencia de los eventos $A_1, A_2, \ldots, A_n$, es

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$$

para cualquier colección de índices distintos $1 \leq i_1 < \cdots < i_k \leq n$ y $2 \leq k \leq n$.

Ejemplo 1.5.2. Considere el espacio de probabilidad uniforme continuo en el cuadrado unitario

$$\Omega = \left\{(x, y) \in \mathbf{R}^2 : 0 < x, y < 1\right\}.$$

Sean los eventos

$$\begin{aligned}
A &= \left\{(x, y) \in \Omega : 0 < x < \frac{1}{2},\ 0 < y < 1\right\}, \\
B &= \left\{(x, y) \in \Omega : 0 < x < 1,\ 0 < y < \frac{1}{4}\right\}, \\
C &= \{(x, y) \in \Omega : 0 < x < 1, 0 < x + 2y < 1\}.
\end{aligned}$$

Entonces

$$\begin{aligned}
A \cap B &= \left\{(x, y) \in \Omega : 0 < x < \frac{1}{2},\ 0 < y < \frac{1}{4}\right\}, \\
A \cap C &= \left\{(x, y) \in \Omega : 0 < x < \frac{1}{2},\ 0 < x + 2y < 1\right\}.
\end{aligned}$$

En la Figura 1.6 se aprecia la gráfica del espacio muestral Ω, así como de los eventos A, B, C, $A \cap B$ y $A \cap C$. Los eventos A y B son independientes, pues

$$P(A \cap B) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = P(A)P(B).$$

Note que el evento B parte al evento A en una relación 1 a 3. Con la misma proporción, el evento B parte al evento A^c. Recíprocamente, el evento A parte al evento B en una relación 1 a 1. Con la misma proporción, el evento A parte al evento B^c. Por otra parte, los eventos A y C son dependientes, pues

$$P(A \cap C) = \frac{0.5 + 0.25}{2} \times 0.5 = 0.1875 \neq 0.125 = 0.5 \times \frac{1 \times 0.5}{2} = P(A)P(C).$$

La Figura 1.6[derecha] muestra la gráfica de dichos eventos.

El siguiente resultado afirma que la independencia entre dos eventos se trasmite a cualquiera de sus respectivos complementos.

Teorema 1.5.3. Sean A y B dos eventos. Entonces

$$A \perp B \quad \Leftrightarrow \quad A \perp B^c \quad \Leftrightarrow \quad A^c \perp B \quad \Leftrightarrow \quad A^c \perp B^c.$$

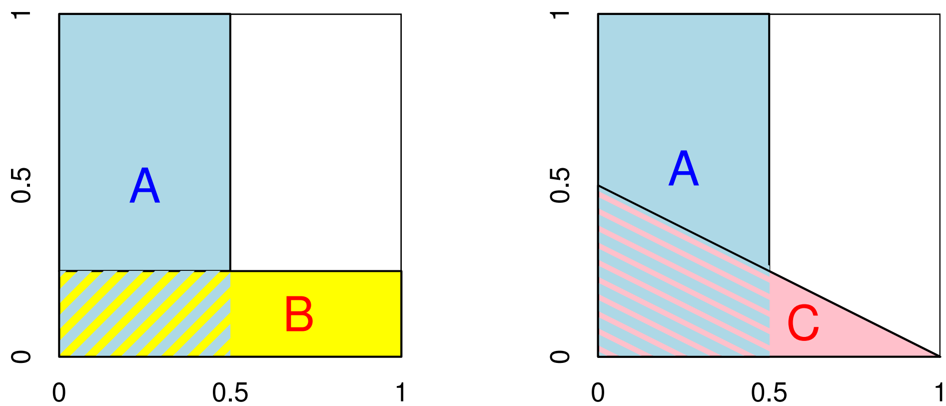


Figura 1.6. [izquierda] Gráfica de dos eventos independientes A y B, en el espacio de muestral $\Omega = (0,1) \times (0,1)$. [derecha] Gráfica de dos eventos dependientes A y C; Ejemplo 1.5.2.

Demostración. Es suficiente verificar la primera implicación. Considere la partición del evento A en dos eventos excluyentes en términos de B y B^c:

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

De la ley de probabilidad total (1.18), se tiene

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

Si $A \perp B$, entonces

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

Por lo que

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A)P(B) + P(A \cap B^c), \end{aligned}$$

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

Se concluye que $A \perp B^c$. $\square$

Los eventos A y Ω son independientes. Más general, si $P(B) = 1$ entonces, los eventos A y B son independientes. Para verificar esta afirmación, note que

$$B \subset A \cup B,$$

y

$$1 = P(B) \leq P(A \cup B) \leq 1.$$

Así que

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = P(A) + 1 - 1 = P(A) = P(A)P(B).$$

Así mismo, los eventos A y $\varnothing$ son independientes. Más general, si $P(B) = 0$ entonces, los eventos A y B son independientes. En tal caso, se tiene

$$A \cap B \subset B$$

y

$$0 \leq P(A \cap B) \leq P(B) = 0.$$

Así que

$$P(A \cap B) = 0 = P(A) \times 0 = P(A)P(B).$$

Se concluye que tanto $\varnothing$ como Ω no aportan información sobre cualquier evento A.

1.6. Ensayos de Bernoulli

Considere un experimento de $n \geq 1$ etapas, tiempos o ensayos. En cada ensayo hay sólo dos resultados posibles: éxito o fracaso, gana o pierde, arriba o abajo, izquierda o derecha, falla o no falla, no falla o falla, se alivia o no, nace o no nace, muere o no muere, si o no, blanco o negro, 1 ó -1, 1 ó 2, 1 ó 0. Asuma además que los ensayos se realizan bajo las mismas condiciones y sus resultados son independientes. Este tipo de experimentos se conoce como *ensayos de Bernoulli*. El éxito se asocia usualmente al número 1 mientras que 0 representa al $fracaso$. El espacio muestral es

$$\begin{aligned}\Omega &= \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i = F, E, \quad i = 1, \ldots, n\} \\ &= \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i = 0, 1, \quad i = 1, \ldots, n\}.\end{aligned}$$

Un evento simple es un vector de dimensión n con valores 0 ó 1 en sus elementos: $\omega = (\omega_1, \ldots, \omega_n)$. El espacio muestral Ω tiene 2^n eventos simples, trayectorias o ramas, pues se deriva de un muestreo ordenado con reemplazo. La Tabla 1.16 muestra un diagrama de árbol que describe al espacio muestral Ω, con $n = 3$ ensayos de Bernoulli. Cada rama o trayectoria representa un evento simple de Ω. La probabilidad de éxito en cada ensayo es la misma:

$$P(\text{éxito}) = P(E) = p; \quad \text{con} \quad 0 \leq p \leq 1. \tag{1.40}$$

Esto equivale a determinar la probabilidad de su complemento o fracaso:

$$P(\text{fracaso}) = P(F) = P(E^c) = 1 - P(E) = 1 - p; \quad 0 \leq p \leq 1. \tag{1.41}$$

Para el ensayo $i = 1, \ldots, n$, se tiene

$$\begin{aligned}P(\omega_i) &= \begin{cases} p & \omega_i = 1 \\ 1 - p & \omega_i = 0 \end{cases} \\ &= p^{\omega_i}(1 - p)^{1 - \omega_i}; \quad \omega_i = 0, 1.\end{aligned}$$

evento simple

ensayo: 1	2	3	suma de éxitos	probabilidad
E	E	E	3	$P(E)P(E)P(E) = p^3$
E	E	F	2	$P(E)P(E)P(F) = p^2(1-p)$
E	F	E	2	$P(E)P(F)P(E) = p^2(1-p)$
E	F	F	1	$P(E)P(F)P(F) = p(1-p)^2$
F	E	E	2	$P(F)P(E)P(E) = p^2(1-p)$
F	E	F	1	$P(F)P(E)P(F) = p(1-p)^2$
F	F	E	1	$P(F)P(F)P(E) = p(1-p)^2$
F	F	F	0	$P(F)P(F)P(F) = (1-p)^3$

Tabla 1.16. Diagrama de árbol del experimento de $n = 3$ ensayos de Bernoulli, con suma de éxitos y probabilidad de cada rama (evento simple); ejemplos 1.2.4 y los de esta sección.

éxito	$p = P(E)$
águila en cara superior de una moneda al caer	$1/2 = 0.5$
número cuatro en cara superior de un dado al caer	$1/6 = 0.1667$
un televisor defectuoso en mostrador	0.003 (3 de cada mil)
precio al alza de acción de una empresa para mañana	0.5074
ganar con el número 00 en la ruleta	$1/38 = 0.0263$
ganar con los números 0 ó 00 en la ruleta	$2/38 = 0.0526$
ganar con los números rojos en la ruleta	$18/38 = 0.4737$
temperatura al amanecer menor que 0°C	0.08

Tabla 1.17. Ejemplos de éxito de una variable aleatoria de distribución Bernoulli, junto con su probabilidad p o un referente de la misma.

Por el supuesto de independencia, la probabilidad el evento simple $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ es

$$\begin{aligned} P(\omega) &= P(\omega_1, \dots, \omega_n) = P(\omega_1) \cdots P(\omega_n) \\ &= p^{\omega_1}(1-p)^{1-\omega_1} \times \cdots \times p^{\omega_n}(1-p)^{1-\omega_n} = p^{\omega_1+\cdots+\omega_n}(1-p)^{(1-\omega_1)+\cdots+(1-\omega_n)} \\ &= p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i} = p^{\text{suma de éxitos}}(1-p)^{n-\text{suma de éxitos}}. \end{aligned} \tag{1.42}$$

Note que esta probabilidad se determina por la suma de éxitos que involucra. Como consecuencia, dos eventos simples que compartan su suma de éxitos tendrán la misma probabilidad.

La Tabla 1.17 muestra algunos ejemplos de éxito, junto con su probabilidad p; ya sea exacta o una referencia. A continuación se analizan tres ejemplos de ensayos de Bernoulli. Recuerde la independencia de los resultados de dichos ensayos.

Ejemplo 1.6.1. Considere el lanzamiento de una moneda legal n veces. En cada ensayo ocurre águila (a, éxito, E, 1) ó sol (s, fracaso, F, 0), con

$$p = P(E) = P(\text{águila}) = \frac{1}{2}.$$

El Ejemplo 1.2.4 describe un experimento de ensayos de Bernoulli, con $n = 3$ y $p = 0.5$. El espacio muestral Ω tiene 2^n eventos simples, todos equiprobables:

$$P(\omega) = P(\omega_1, \dots, \omega_n) = \frac{1}{2^n},$$

donde $\omega_i = 0, 1$; para $i = 1, \dots, n$. De hecho, por (1.42), se tiene

$$\begin{aligned} P(\omega_1, \dots, \omega_n) &= p^{\text{suma de éxitos}}(1-p)^{n-\text{suma de éxitos}} = \left(\frac{1}{2}\right)^{\text{suma de éxitos}} \left(1 - \frac{1}{2}\right)^{n-\text{suma de éxitos}} \\ &= \left(\frac{1}{2}\right)^{\text{suma de éxitos}} \left(\frac{1}{2}\right)^{n-\text{suma de éxitos}} = \left(\frac{1}{2}\right)^{\text{suma de éxitos}+n-\text{suma de éxitos}} \\ &= \left(\frac{1}{2}\right)^n = \frac{1}{2^n}. \end{aligned}$$

Si en particular $n = 3$, se obtiene

$$\begin{aligned} \Omega &= \{aaa, aas, asa, ass, saa, sas, ssa, sss\}, \\ \#\Omega &= 2^n = 2^3 = 8, \\ P(aaa) &= P(aas) = \cdots = P(sss) = \frac{1}{2^3} = \frac{1}{8}. \end{aligned}$$

Considere los eventos

$$\begin{aligned} &A_1 : \text{águila en primer ensayo}, \\ &A_2 : \text{águila en segundo ensayo}, \\ &A_3 : \text{águila en tercer ensayo}. \end{aligned}$$

Por construcción del espacio de probabilidad, los eventos A_1, A_2 y A_3 son independientes. Lo que se ilustra continuación. Como

$$\begin{aligned} A_1 &= \{aaa, aas, asa, ass\}, \\ A_2 &= \{aaa, aas, saa, sas\}, \\ A_3 &= \{aaa, asa, saa, ssa\}, \\ A_1 \cap A_2 &= \{aaa, aas\}, \\ A_1 \cap A_3 &= \{aaa, asa\}, \\ A_2 \cap A_3 &= \{aaa, saa\} \\ A_1 \cap A_2 \cap A_3 &= \{aaa\}, \end{aligned}$$

entonces

$$\begin{aligned} P(A_1 \cap A_2) &= \frac{2}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2), \\ P(A_1 \cap A_3) &= \frac{2}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_3), \\ P(A_2 \cap A_3) &= \frac{2}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2)P(A_3), \\ P(A_1 \cap A_2 \cap A_3) &= \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2)P(A_3). \end{aligned}$$

Por lo tanto, son independientes los eventos A_1, A_2 y A_3.

Ejemplo 1.6.2. Considere el juego de la ruleta con 38 números: 18 rojos, 18 negros, cero y doble cero; véanse Ejemplo 1.3.1 y Figura 1.4. La estrategia de un jugador es apostar a los rojos en $n = 5$ juegos. La probabilidad de ganar en cada juego es

$$P(E) = p = \frac{18}{38} = 0.4737.$$

El espacio muestral es $\Omega = \{(\omega_1, \ldots, \omega_5) : \omega_i = 0, 1,\ i = 1, \ldots, 5\}$, mientras que su cardinalidad es $\#\Omega = 2^n = 2^5 = 32$. La probabilidad de cada evento simple depende sólo de la suma de éxitos

$$\begin{aligned} P(\omega_1, \ldots, \omega_5) &= [P(E)]^{\text{suma de éxitos}}(1 - P(E))^{5-\text{suma de éxitos}} \\ &= p^{\text{suma de éxitos}}(1-p)^{5-\text{suma de éxitos}} \\ &= 0.4737^{\text{suma de éxitos}} \times 0.5263^{5-\text{suma de éxitos}}. \end{aligned}$$

Como los eventos simples no son equiprobables, entonces no aplica la fórmula de probabilidad (1.13). Tanto el jugador como la casa tienen interés en los siguientes eventos:

$$\begin{aligned} &A_0 : \text{ganar 0 veces (perder todos los juegos)}, \\ &A_1 : \text{ganar 1 vez}, \\ &A_2 : \text{ganar 2 veces}, \\ &A_3 : \text{ganar 3 veces}, \\ &A_4 : \text{ganar 4 veces}, \\ &A_5 : \text{ganar 5 veces (ganar todos los juegos)}. \end{aligned}$$

La Tabla 1.18 describe los eventos simples que integran a dichos eventos de interés. Esta tabla resume también sus cardinalidades y probabilidades. Note que $A_0, \ldots, A_5$ forman una partición del espacio muestral Ω, pues son eventos excluyentes y

$$\Omega = A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5.$$

El evento A_0 es simple. Le corresponde el único elemento de Ω de 5 fracasos en 5 ensayos:

$$A_0 = \{FFFFF\}.$$

Por el supuesto de independencia, su probabilidad es

$$\begin{aligned} P(A_0) &= P(FFFFF) = P(F)P(F)P(F)P(F)P(F) = [P(F)]^5 \\ &= (1-p)^5 = 0.5263^5 = 0.0404 = 4.04\,\%. \end{aligned}$$

Note que

$$P(A_0) = \binom{5}{0} p^0 (1-p)^{5-0} = \binom{5}{0} 0.4737^0 \times 0.5263^{5-0}.$$

El evento A_1 se compone de 5 eventos simples:

$$A_1 = \{EFFFF, FEFFF, FFEFF, FFFEF, FFFFE\}.$$

Como estos eventos simples son equiprobables, entonces

$$\begin{aligned} P(A_1) &= 5 \times P(EFFFF) = 5 \times P(E)P(F)P(F)P(F)P(F) \\ &= 5 \times 0.4737 \times 0.5263^4 = 0.1817 = 18.17\,\%. \end{aligned}$$

Note que la letra E se posiciona de $\binom{5}{1} = 5$ maneras diferentes. Por lo que

$$P(A_1) = \binom{5}{1} p^1 (1-p)^{5-1} = \binom{5}{1} \times 0.4737^1 \times 0.5263^{5-1}.$$

La probabilidad del evento A_2 es

$$\begin{aligned} P(A_2) &= 10P(EEFFF) = \binom{5}{2} P(E)P(E)P(F)P(F)P(F) \\ &= \binom{5}{2} p^2 (1-p)^{5-2} = 10 \times 0.4737^2 \times 0.5263^3 = 0.3271 = 32.71\,\%. \end{aligned}$$

Aquí las dos letras EE se posicionan de $\binom{5}{2}$ diferentes maneras. De la misma manera, se obtienen las probabilidades de los eventos A_3, A_4 y A_5:

$$\begin{aligned} P(A_3) &= 10P(EEEFF) = \binom{5}{3} P(E)P(E)P(E)P(F)P(F) \\ &= \binom{5}{3} p^3 (1-p)^{5-3} = 10 \times 0.4737^3 \times 0.5263^2 = 0.2944 = 29.44\,\%, \\ P(A_4) &= 5P(EEEEF) = \binom{5}{4} P(E)P(E)P(E)P(E)P(F) \\ &= \binom{5}{4} p^4 (1-p)^{5-4} = 5 \times 0.4737^4 \times 0.5263^1 = 0.1325 = 13.25\,\%, \\ P(A_5) &= P(EEEEE) = \binom{5}{5} P(E)P(E)P(E)P(E)P(E) \\ &= \binom{5}{5} p^5 (1-p)^{5-5} = 0.4737^5 \times 0.5263^0 = 0.0239 = 2.39\,\%. \end{aligned}$$

evento	número de rojos (éxitos)	cardinalidad	probabilidad
$FFFFF$	0	$\binom{5}{0} = 1$	0.0404
$EFFFF, FEFFF,$ $FFEFF, FFFEF, FFFFE$	1	$\binom{5}{1} = 5$	0.1817
$EEFFF, EFEFF, EFFEF,$ $EFFFE, FEEFF, FEFEF,$ $FEFFE, FFEEF, FFEFE, FFFEE$	2	$\binom{5}{2} = 10$	0.3271

$EEEFF, EEFEF, EEFFE, EFEEF,$ $EFEFE, EFFEE, FEEEF,$ $FEEFE, FEFEE, FFEEE$	3	$\binom{5}{3} = 10$	0.2944
$EEEEF, EEEFE, EEFEE,$ $EFEEE, FEEEE$	4	$\binom{5}{4} = 5$	0.1325
$EEEEE$	5	$\binom{5}{5} = 1$	0.0239
	suma	$2^5 = 32$	1

Tabla 1.18. Eventos en términos del número de rojos (éxitos) en $n = 5$ juegos de la ruleta, junto con sus respectivas cardinalidades y probabilidades; Ejemplo 1.6.2.

De la Tabla 1.18, destaca lo siguiente.

1. La secuencia de coeficientes binomiales que acompaña a cada probabilidad es

$$\binom{5}{0}, \binom{5}{1}, \binom{5}{2}, \binom{5}{3}, \binom{5}{4}, \binom{5}{5},$$

donde $\binom{5}{k}$ representa el número de posiciones seleccionadas por combinación de las k letras E en $n = 5$ ensayos; para $k = 0, 1, 2, 3, 4, 5$.

2. La siguiente identidad representa al total de eventos simples del experimento.

$$\#\Omega = 2^5 = \binom{5}{0} + \binom{5}{1} + \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5}.$$

Esta expresión se deduce también con la fórmula del binomio; véase (20) del Apéndice.

3. La estrategia del jugador lo lleva al triunfo si ocurre la unión excluyente de eventos $A_3 \cup A_4 \cup A_5$. Su probabilidad es

$$\begin{aligned} P(A_3 \cup A_4 \cup A_5) &= P(A_3) + P(A_4) + P(A_5) \\ &= 0.2944 + 0.1325 + 0.0239 = 0.4508. \end{aligned}$$

Compare esta probabilidad con la de ganar en sólo un juego, que es 0.4737. Por lo tanto, se sugiere jugar sólo una vez en lugar de la estrategia de cinco ensayos. ¿Conviene jugar más de una vez?, ¿será jugar sólo una vez la mejor estrategia?

Ejemplo 1.6.3. Control de calidad aplicado a la industria del concreto. Considere un lote de $n = 7$ cilindros de ensayo de concreto. Cada cilindro se somete a una prueba destructiva de resistencia a la comprensión. La Figura 1.7 muestra la compresión de un cilindro de concreto. La fuerza de resistencia aumenta paulatinamente hasta que se fracture el cilindro. Este pasa la prueba de resistencia (fracaso) si supera un umbral de fuerza. En caso contrario, el cilindro no pasa la prueba (éxito). El espacio muestral es

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_7) : \omega_i = 0, 1, \; i = 1, \ldots, 7\}.$$

Un evento simple es un vector de dimensión $n = 7$, cuyos elementos son cero o uno. Por el Ejercicio 1.21, el total de eventos simples del espacio muestral es $\#\Omega = 2^n = 2^7 = 128$. Algunos ejemplos de eventos simples de este experimento son los siguientes:

$$FFFFFFF,\ FEFFFFF,\ FFEFFFF,\ FFFFEFE,\ FFFEEE.$$

Asuma que la probabilidad de éxito (falla del cilindro) es la misma para cada ensayo:

$$p = P(E) = P(falla) = 0.003 = 0.3\,\% \quad \text{(3 de cada 1000)}.$$

Por el supuesto de independencia, la probabilidad de un evento simple es el producto de las respectivas probabilidades de cada ensayo:

$$P(\omega_1, \ldots, \omega_7) = P(\omega_1) \cdots P(\omega_7),$$

donde

$$\begin{aligned} P(\omega_i) &= \begin{cases} 1-p & \omega_i = 0 \\ p & \omega_i = 1 \end{cases} \\ &= \begin{cases} 0.997 & \omega_i = 0 \\ 0.003 & \omega_i = 1 \end{cases} \\ &= 0.003^{\omega_i}(1-0.003)^{1-\omega_i} \\ &= 0.003^{\omega_i} 0.997^{1-\omega_i}; \end{aligned}$$

para $i = 1, \ldots, 7$. Así

$$\begin{aligned} P((w_1, \ldots, w_7)) &= p^{\text{suma de éxitos}}(1-p)^{7-\text{suma de éxitos}} \\ &= 0.003^{\text{suma de éxitos}} \times 0.997^{7-\text{suma de éxitos}}, \end{aligned}$$

donde suma de éxitos $= \sum_{i=1}^{7} \omega_i$. En particular, un evento de interés es

$$\begin{aligned} B_0 &: \text{no hay cilindros con falla} \\ &: \text{hay 7 fracasos} \\ &= \{FFFFFFF\}. \end{aligned}$$

Su probabilidad es

$$P(B_0) = P(FFFFFFF) = \binom{7}{0} p^0 (1-p)^{7-0} = 0.003^0 \times 0.997^7 = 0.9792 = 97.92\,\%.$$

Este evento simple tiene un gran peso en el espacio muestral Ω; que se compone de 128 eventos simples. Su complemento, de 127 eventos simples, tiene probabilidad $0.0208 = 2.08\,\%$. Ello se debe a que la probabilidad de fracaso (no falla del cilindro) es muy alta. Como los eventos simples de Ω no son equiprobables, la fórmula (1.13) no aplica para este espacio de probabilidad. Ahora considere el evento

$$\begin{aligned} B_1 &: \text{hay sólo un éxito en 7 ensayos (sólo un cilindro con falla)} \\ &= \{EFFFFFF,\ FEFFFFF,\ FFEFFFF,\ FFFEFFF, \\ &\quad FFFFEFF,\ FFFFFEF,\ FFFFFFE\}. \end{aligned}$$

Los siete eventos simples de este conjunto son equiprobables. Por lo que

$$\begin{aligned} P(B_1) &= 7 \times P(EFFFFFF) = 7 \times P(E)[P(F)]^6 \\ &= \binom{7}{1} p^1 (1-p)^{7-1} = 7 \times 0.003^1 \times 0.997^6 \\ &= 0.0206 = 2.06\,\%. \end{aligned}$$

De forma similar, se obtiene

$$\begin{aligned} P(B_2) &= P(2 \text{ éxitos}) = \binom{7}{2} p^2 (1-p)^5 = 21 \times 0.003^2 \times 0.997^5 = 0.0002, \\ P(B_3) &= P(3 \text{ éxitos}) = \binom{7}{3} p^3 (1-p)^4 = 35 \times 0.003^3 \times 0.997^4 = 9.34 \times 10^{-7}, \\ P(B_4) &= P(4 \text{ éxitos}) = \binom{7}{4} p^4 (1-p)^3 = 35 \times 0.003^4 \times 0.997^3 = 2.81 \times 10^{-9}, \\ P(B_5) &= P(5 \text{ éxitos}) = \binom{7}{5} p^5 (1-p)^2 = 21 \times 0.003^5 \times 0.997^2 = 5.07 \times 10^{-12}, \\ P(B_6) &= P(6 \text{ éxitos}) = \binom{7}{6} p^6 (1-p)^1 = 7 \times 0.003^6 \times 0.997^1 = 5.09 \times 10^{-15}, \\ P(B_7) &= P(7 \text{ éxitos}) = \binom{7}{7} p^7 (1-p)^0 = 0.003^7 \times 0.997^0 = 2.19 \times 10^{-18}. \end{aligned}$$

Figura 1.7. Cilindro de ensayo de concreto en una prueba de compresión; Ejemplo 1.6.3.

1.7. Probabilidad condicional

En esta sección se desarrollará la probabilidad condicional.

La probabilidad condicional se refiere a la probabilidad relativa a información parcial o previa disponible del fenómeno de estudio. Por ejemplo, con registros históricos se tiene una estimación de la probabilidad de que llueva el próximo 24 de junio. Dicha probabilidad cambia si un día anterior llovió. El cambio de la probabilidad implica que los dos eventos involucrados son dependientes (no independientes); llueve el 24 de junio y el día anterior. Por otro lado, se pueden considerar independientes los eventos: llueve el próximo 24 de junio o el mismo día del siguiente año. La distancia temporal otorga independencia entre ambos eventos.

Cuando dos eventos A y B son independientes, aplica la fórmula (1.34) para el cálculo de la probabilidad de la intersección. Sin embargo, si los eventos son dependientes, aplica una fórmula similar en términos de la probabilidad condicional.

Cuando en un experimento se sabe que ocurrió un evento A, entonces la incertidumbre cambia. Antes del experimento se sabe que ocurrirá cierto evento simple $\omega \in \Omega$. Ahora lo relevante es que tal evento simple está en A. El conjunto de resultados posibles del experimento se reduce de Ω al evento A. Ante el nuevo escenario, todos los eventos de Ω pueden sufrir de algún cambio en sus probabilidades. Por ejemplo, los subconjuntos de A ahora tienen más posibilidad de ocurrir. En cambio, el evento A^c jamás ocurrirá; a pesar de que $P(A^c) > 0$.

Definición 1.7.1. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Dado un evento A, con $P(A) > 0$, la *probabilidad condicional* respecto de A es la medida definida en $\mathcal{A}$ como

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}; \quad B \in \mathcal{A}. \tag{1.43}$$

Si $P(A) = 0$, la probabilidad condicional es indefinida.

La probabilidad condicional $P(\cdot \mid A)$ es otra medida de probabilidad definida en $\mathcal{A}$. Esta medida es comparable con la medida *incondicional* $P(\cdot)$. Note que

$$P(A \cap B) = P(A)P(B \mid A); \quad \text{siempre que} \quad P(A) > 0, \tag{1.44}$$

$$P(A \cap B) = P(A \mid B)P(B); \quad \text{siempre que} \quad P(B) > 0. \tag{1.45}$$

Si $P(A) > 0$ o $P(B) > 0$, entonces (1.44) o (1.45) son fórmulas para el cálculo de la probabilidad del evento intersección $A \cup B$. Si en particular son independientes los eventos A y B, por (1.34), debe cumplirse

$$\begin{aligned} P(B \mid A) = P(B); &\quad \text{si} \quad P(A) > 0, \\ P(A \cap B) = P(A); &\quad \text{si} \quad P(B) > 0. \end{aligned}$$

Cabe recordar que si $P(A) = 0$ o $P(B) = 0$, entonces no está definida la probabilidad condicional $P(A \mid B)$ o $P(A \mid B)$; respectivamente. Sin embargo, en ambos casos, se tiene la independencia de A y B. El siguiente teorema formaliza esta caracterización de independencia.

Teorema 1.7.2. Sean A y B dos eventos tal que $P(A) > 0$ y $P(B) > 0$. Entonces, se cumplen las siguientes equivalencias:

$$\begin{aligned} A \perp B \quad &\Leftrightarrow \quad A \text{ y } B \text{ son eventos independientes} \\ &\Leftrightarrow \quad P(A \cap B) = P(A)P(B) \\ &\Leftrightarrow \quad P(B \mid A) = P(B) \\ &\Leftrightarrow \quad P(A \mid B) = P(A). \end{aligned}$$

Demostración. Considere dos eventos A y B con $P(A) > 0$ y $P(B) > 0$. Es suficiente verificar la equivalencia

$$P(A \cap B) = P(A)P(B) \quad \Leftrightarrow \quad P(B \mid A) = P(B).$$

$\underline{(\Rightarrow)}$ Si ambos eventos son independientes, entonces

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

Así que

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

$\underline{(\Leftarrow)}$ El recíproco se obtiene al proceder en sentido contrario. Como $P(B \mid A) = P(B)$, entonces

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

□

Salvo el caso de independencia, son diferentes las probabilidades condicional e incondicional, como se describe en el siguiente ejemplo.

Ejemplo 1.7.3 (Urna de Polya)**.** Considere una caja con r bolas rojas y b bolas negras. Se extrae al azar una bola y se regresa a la caja luego de registrar su color. Agregue a la caja $c > 0$ bolas del color de la bola seleccionada. Calcule la probabilidad de los siguientes eventos

$$\begin{aligned} &R_1 : \text{primera bola es roja}, \qquad &&B_1 : \text{primera bola es negra}, \\ &R_2 : \text{segunda bola es roja}, &&B_2 : \text{segunda bola es negra}. \end{aligned}$$

Note que $R_1^c = B_1$ y $R_2^c = B_2$. La Tabla 1.19 muestra el diagrama de árbol del espacio muestral de este experimento de dos etapas, junto con la probabilidad de los eventos simples. Se calcularán tres probabilidades del evento R_2: $P(R_2)$, $P(R_2 \mid R_1)$ y $P(R_2 \mid B_1)$. Por la ley de probabilidad total (1.18) y (1.44), se tiene

$$\begin{aligned} P(R_2) &= P(R_2 \cap \Omega) = P(R_2 \cap (R_1 \cup B_1)) \\ &= P(R_1 \cap R_2) + P(B_1 \cap R_2) \\ &= P(R_1)P(R_2 \mid R_1) + P(B_1)P(R_2 \mid B_1), \end{aligned}$$

con

$$\begin{aligned} P(R_1) &= \frac{r}{b+r}, \qquad & P(R_2 \mid R_1) &= \frac{r+c}{b+r+c}, \\ P(B_1) &= \frac{b}{b+r}, & P(R_2 \mid B_1) &= \frac{r}{b+r+c}. \end{aligned}$$

Entonces

$$
\begin{aligned}
P(R_2) &= \frac{r}{b+r}\,\frac{r+c}{b+r+c} + \frac{b}{b+r}\,\frac{r}{b+r+c} = \frac{r}{b+r}\,\frac{r+c}{b+r+c} \\
&= \frac{r}{b+r}\,\frac{b+r+c}{b+r+c} = \frac{r}{b+r} = P(R_1), \\
P(B_2) &= 1 - P(B_2^c) = 1 - P(R_2) = 1 - \frac{r}{b+r} = \frac{b}{b+r} = P(B_1).
\end{aligned}
$$

Los eventos R_1 y R_2 tienen la misma probabilidad mas no son independientes. Si ocurre R_1 entonces, aumentan las posibilidades de R_2. Por ejemplo, si $r = 10$ y $b = c = 5$, entonces

$$
P(R_2 \mid R_1) = \frac{r+c}{b+r+c} = \frac{10+5}{15+5} = \frac{15}{20} = \frac{3}{4} > \frac{2}{3} = \frac{10}{15} = \frac{r}{b+r} = P(R_2).
$$

En cambio, dado B_1, disminuye la probabilidad de R_2:

$$
P(R_2 \mid B_1) = \frac{10}{20} = \frac{1}{2} < \frac{2}{3} = P(R_2).
$$

Se concluye

$$
P(R_2 \mid B_1) < P(R_2) < P(R_2 \mid R_1).
$$

Cabe mencionar que la probabilidad condicional tiene una mayor capacidad de predicción del futuro que su contraparte incondicional. De manera análoga, el pasado se explica con el presente observado (retróstico). Véase Ejercicio 1.24, para verificar las siguientes desigualdades.

$$
P(R_1 \mid B_2) < P(R_1) < P(R_1 \mid R_2).
$$

evento simple, etapa 1	evento simple, etapa 2	probabilidad
r	$r+c$	$P(R_1 \cap R_2) = P(R_1)P(R_2 \mid R_1) = \frac{r}{b+r}\,\frac{r+c}{b+r+c}$
r	b	$P(R_1 \cap B_2) = P(R_1)P(B_2 \mid R_1) = \frac{r}{b+r}\,\frac{b}{b+r+c}$
b	r	$P(B_1 \cap R_2) = P(B_1)P(R_2 \mid B_1) = \frac{b}{b+r}\,\frac{r}{b+r+c}$
b	$b+c$	$P(B_1 \cap B_2) = P(B_1)P(B_2 \mid B_1) = \frac{b}{b+r}\,\frac{b+c}{b+r+c}$

Tabla 1.19. Diagrama de árbol de la urna de Polya de $n = 2$ etapas, con las probabilidades de los eventos simples; Ejemplo 1.7.3.

Ejemplo 1.7.4. Considere el juego de cartas del Ejemplo 1.4.8, en donde se selecciona una mano al azar: 5 cartas de las 52 disponibles.

1. Sean los eventos

$$A : \text{mano con al menos un par de reinas,}$$
$$B : \text{tercia de reinas.}$$

Note que $B \subset A$. Entonces

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

Por (1.33), se tiene $\#B = \binom{4}{3} \times \binom{12}{2} 4 \times 4 = 4\,224$. Para el evento A, considere su partición en tres eventos excluyentes

$$A = A_2 \cup A_3 \cup A_4,$$

con

$$A_2 : \text{mano con exactamente 2 reinas,}$$
$$A_3 : \text{mano con exactamente 3 reinas,}$$
$$A_4 : \text{mano con exactamente 4 reinas.}$$

Note que el evento A_2 incluye a las manos: par de reinas, dos pares con un par de reinas y full con un par de reinas. El evento A_3 integra a las manos: tercia de reinas (el evento B) y full de reinas. El póquer de reinas es representado por A_4. La cardinalidad del evento A se obtiene con el muestreo hipergeométrico:

$$\#A = \binom{4}{2}\binom{48}{3} + \binom{4}{3}\binom{48}{2} + \binom{4}{4}\binom{48}{1} = 108\,336.$$

Por lo que

$$P(B \mid A) = \frac{4224}{108336} = 0.0389 = 3.89\,\%.$$

En un juego real, ¿cómo aplicaría usted los resultados obtenidos de las probabilidades condicional e incondicional?

2. Considere una mano que tiene al menos dos reyes y exactamente dos ochos: $KK88z$, con $z \neq 8$. Obtenga la probabilidad condicional de un full de reyes y ochos $KKK88$. Defina los eventos

$$C : \text{mano full de reyes y ochos; } KKK88,$$
$$D : \text{mano con al menos dos reyes y dos ochos; } KK88z,\ z \neq 8.$$

Note que $C \subset D$. Además, el evento D se representa con la unión de los siguientes dos eventos excluyentes:

$$D = (D \cap C) \cup (D \cap C^c) = C \cup (D \cap C^c),$$

donde

$$D \cap C^c : \text{mano par de reyes y ochos; } KK88z,\ z \neq K, 8.$$

Así que

$$P(C \mid D) = \frac{P(C \cap D)}{P(D)} = \frac{P(C)}{P(D)},$$

con

$$P(D) = P(C) + P(D \cap C^c) = \frac{\binom{4}{3}\binom{4}{2}\binom{44}{0}}{\binom{52}{5}} + \frac{\binom{4}{2}\binom{4}{2}\binom{44}{1}}{\binom{52}{5}}.$$

Por lo tanto

$$P(C \mid D) = \frac{\binom{4}{3}\binom{4}{2}}{\binom{4}{3}\binom{4}{2} + \binom{4}{2}\binom{4}{2}44} = \frac{24}{24 + 36 \times 44} = 0.0149 = 1.49\,\%.$$

1.7.1. Regla de Bayes

En esta sección se desarrolla la fórmula de Bayes. Se describirá la regla de Bayes mediante dos ejemplos. El primero es aplicado a la medición de la eficiencia de una prueba médica. El segundo ejemplo corresponde a la evaluación del detector de mentiras.

Ejemplo 1.7.5. Considere una población en donde el $0.5\,\%$ de sus individuos tiene cierta enfermedad; *prevalencia* de 5 por cada 1000 individuos. Una prueba de laboratorio tiene $95\,\%$ de efectividad para detectar la enfermedad cuando realmente está presente. En cambio, la prueba reporta como falso positivo al $1\,\%$ de las personas sanas. ¿cuál es la probabilidad de que una persona diagnosticada como enferma, realmente lo esté? Para contestar la pregunta, construya una tabla de *contingencia* de 2×2, con los cuatro resultados posibles del experimento; como se muestra en la Tabla 1.20. La lectura de las probabilidades condicionales es por renglón.

Sean los eventos

$$\begin{aligned} E &= \text{paciente está enfermo}, \\ D &= \text{diagnóstico es positivo}. \end{aligned}$$

La probabilidad de interés es $P(E \mid D)$. Los datos de este problema se resumen en las siguientes tres probabilidades

$$P(E) = 0.005, \quad P(D \mid E) = 0.95, \quad P(D \mid E^c) = 0.01.$$

enfermedad		resultado de la prueba	
	$100\,\%$	negativo	positivo
no presente	$99.5\,\%$	$99\,\%$	*falso positivo* $1\,\%$
presente	$0.5\,\%$	*falso negativo* $5\,\%$	$95\,\%$

Tabla 1.20. Tabla de contingencia de 2×2, de una prueba de diagnóstico médico.

Por definición de probabilidad condicional, se tiene

$$P(E \mid D) = \frac{P(E \cap D)}{P(D)}.$$

Pero como

$$P(E \cap D) = P(E)P(D \mid E),$$

entonces

$$P(E \mid D) = \frac{P(E)P(D \mid E)}{P(D)}.$$

Así mismo, recuerde que el conjunto D se parte en dos eventos excluyentes:

$$D = (D \cap E) \cup (D \cap E^c).$$

Por la ley de probabilidad total (1.18), se tiene

$$\begin{aligned} P(D) &= P(D \cap E) + P(D \cap E^c) \\ &= P(E)P(D \mid E) + P(E^c)P(D \mid E^c) \end{aligned}$$

y

$$P(E \mid D) = \frac{P(D \mid E)P(E)}{P(E)P(D \mid E) + P(E^c)P(D \mid E^c)}.$$

De esta manera

$$\begin{aligned} P(E \mid D) &= \frac{0.005 \times 0.95}{0.005 \times 0.95 + 0.995 \times 0.01} = \frac{0.00475}{0.00475 + 0.00995} \\ &= \frac{0.00475}{0.0147} = 0.3231 = 32.31\,\%. \end{aligned}$$

Si bien la prueba detecta al 95 % de las personas enfermas, el 32.31 % de las personas positivamente diagnosticadas realmente tendrán la enfermedad. Este paradójico hecho pone en duda la calidad de la prueba. su debilidad se debe a la escasa prevalencia de la enfermedad en la población objetivo. De hecho, la calidad de la prueba mejora cuando la enfermedad tiene una mayor prevalencia. Si se considera que el grupo de personas que se realiza la prueba es a quienes cuentan con un síntoma, como fiebre, enrojecimiento de la piel, entre otros. También, es posible que paciente cuente con un pre diagnóstico. Ante ello, asuma que la prevalencia de la enfermedad es del 10 %. Entonces $P(E) = 0.1$ y

$$\begin{aligned} P(E \mid D) &= \frac{P(E)P(D \mid E)}{P(E)P(D \mid E) + P(E^c)P(D \mid E^c)} \\ &= \frac{0.1 \times 0.95}{0.1 \times 0.95 + 0.9 \times 0.01} \\ &= 0.9135 = 91.35\,\%. \end{aligned}$$

Por otro lado, este resultado cuantifica también el riesgo de la enfermedad en el estrato con pre diagnóstico respecto de la población general:

$$\frac{P(E \mid D)}{P(E)} = \frac{0.3231}{0.005} = 64.626.$$

El riesgo de enfermedad de un individuo con diagnóstico positivo es 64 veces mayor al de uno de la población general. En cambio, el riesgo disminuye si la prueba resulta negativa. Se deja lector la verificación de la última afirmación con el cálculo del cociente de probabilidades $P(E \mid D^c)/P(E)$.

Ejemplo 1.7.6 (Fiabilidad del polígrafo). *Muchos científicos de todo el mundo se han rebelado contra la utilización del polígrafo por los organismos de seguridad y las entidades privadas. Un informe de la Academia Nacional de Ciencias de los Estados Unidos, realizado sin la colaboración de expertos en su uso -según denuncia de la American Polygraph Association (APA), un lobby de la industria del polígrafo-, estimó que de aplicarse a una muestra a* 10 000 *personas, de las cuales diez fueran espías, el resultado más probable es que solo ocho de éstas serían identificados por el polígrafo como tales, a la vez que serían considerados probables espías y rechazados* 1598 *(casi el* 16 % *de los inocentes). Otra investigación, de la Asociación Estadounidense de Psicología, concluyó que los resultados del polígrafo como detector de mentiras no se han distinguido del efecto placebo*; [15]. Estratifique a la población como espías E y no espías E^c. Denote como D al evento de diagnosticar como espía a una persona. Las características del detector de mentiras se resumen como sigue:

$$
\begin{aligned}
P(\quad) &= \frac{10}{10\,000} = 0.001 = 0.1\,\%, \\
P(\quad \mid \quad) &= \frac{8}{10} = 0.8 = 80\,\%, \\
P(\quad \mid \quad\;) &= \frac{1598}{9\,990} = 0.16 = 16\,\%.
\end{aligned}
$$

El evento de mayor interés es $P(E \mid D)$:

$$
\begin{aligned}
P(E \mid D) &= \frac{P(E)P(D \mid E)}{P(E)P(D \mid E) + P(E^c)P(D \mid E^c)} \\
&= \frac{0.001 \times 0.8}{0.001 \times 0.8 + 0.999 \times 0.16} \\
&= 0.0049 = 0.49\,\%.
\end{aligned}
$$

Esta es una probabilidad muy baja. En primera instancia, se concluye que un resultado positivo del polígrafo no es una prueba suficiente para incriminar al sospechoso, el cual debe quedar libre. Sin embargo, la APA no estuvo de acuerdo, alegando que el número de espías relativo a la muestra subestima a los que realmente hay. Con lo cual, en la práctica, $P(E)$ debe ser mayor que 0.001. En efecto, con el diseño muestral propuesto, se estiman las probabilidades $P(D \mid E)$ y $P(D \mid E^c)$, mas no $P(E)$. Si bien, como argumenta la APA, esta última probabilidad no es fácil de estimar, al conceder una probabilidad mayor del evento E, la confiabilidad del polígrafo no mejora significativamente. Por ejemplo, si

$$P(E) = 0.01 = 1\,\%,$$

entonces

$$\begin{aligned} P(E \mid D) &= \frac{P(D \mid E)P(E)}{P(D \mid E)P(E) + P(D \mid E^c)P(E^c)} \\ &= \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.16 \times 0.99} \\ &= 0.0481 = 4.81\,\%. \end{aligned}$$

La probabilidad del evento de interés aumentó casi diez veces más. Pero esta es aún insuficiente para considerar al detector de mentiras como confiable. Por lo tanto, se concluye que no es confiable el resultado de una prueba positiva del polígrafo. Por último, en el pasado reciente de México, ha ocurrido caso graves de violación de los derechos humanos en la impartición de justicia. El Gobierno Federal del año 2008 compró un detector eléctrico de drogas, armas, dinero y explosivos. Mediante la regla de Bayes, debió probarse la confiabilidad de este aparato. Por el contrario, La implementación del detector envió a prisión a más de 5 mil personas. Luego se comprobó que solo encendían sus indicadores en forma aleatoria, similar al del semáforo aduanal; [5].

La fórmula de Bayes es

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)} = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A^c)P(B \mid A^c)}. \tag{1.46}$$

Para el denominador se usó la partición del espacio muestral Ω, según los eventos A y A^c, junto con la ley de probabilidad total (1.18):

$$B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c) = (A \cap B) \cup (A^c \cap B)$$

y

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A)P(B \mid A) + P(A^c)P(B \mid A^c). \end{aligned} \tag{1.47}$$

La fórmula de Bayes se generaliza de la siguiente manera. Sea $A_1, A_2, \ldots, A_k$ una partición del espacio muestral Ω, es decir, una colección de eventos excluyentes cuya unión es Ω:

$$\bigcup_{i=1}^{k} A_i = \Omega \quad \text{y} \quad A_i \cap A_j = \varnothing; \quad \text{para} \quad i \neq j, i, j = 1, \ldots, k.$$

De la ley de probabilidad total (1.19), se tiene

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_k \cap B) \\ &= P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_k)P(B \mid A_k). \end{aligned} \tag{1.48}$$

Aquí se usó

$$\begin{aligned} B &= B \cap \Omega = B \cap (A_1 \cup \cdots \cup A_k) \\ &= (B \cap A_1) \cup \cdots \cup (B \cap A_k) \\ &= (A_1 \cap B) \cup \cdots \cup (A_k \cap B). \end{aligned}$$

La regla de Bayes resulta

$$\begin{aligned} P(A_1 \mid B) &= \frac{P(B \mid A_1)P(A_1)}{P(B)} \\ &= \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + \cdots + P(A_k)P(B \mid A_k)}. \end{aligned} \tag{1.49}$$

De la misma, manera, para $i = 1, \ldots, k$, se obtiene

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \cdots + P(A_k)P(B \mid A_k)}.$$

Recuerde

$$P(\Omega) = P(A_1) + P(A_2) + \cdots + P(A_k) = 1.$$

1.8. Ejercicios

Esta es la sección de ejercicios del capítulo de espacios de probabilidad.

1. Pruebe formalmente las siguientes relaciones

$$(A \cap B) \subset A \subset (A \cup B).$$

2. Pruebe formalmente la propiedad conmutativa de la intersección

$$A \cap B = B \cap A.$$

3. Pruebe formalmente las propiedades asociativas:

$$A \cup (B \cup C) = (A \cup B) \cup C = (A \cup C) \cup B = A \cup B \cup C$$

y

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

Nota: es suficiente verificar la primera igualdad de ambas expresiones.

4. Pruebe formalmente la siguiente propiedad distributiva

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

5. Pruebe formalmente la propiedad

$$(A^c)^c = A.$$

6. Con la propiedad distributiva, pruebe las siguientes dos igualdades de conjuntos:

$$(A \cup B) \cap (C \cup D) = (A \cap C) \cup (A \cap D) \cup (B \cap C) \cup (B \cap D)$$

y

$$(A \cap B) \cup (C \cap D) = (A \cup C) \cap (A \cup D) \cap (B \cup C) \cap (B \cup D).$$

7. Pruebe formalmente la ley de De Morgan (1.3):

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

8. Pruebe formalmente la versión infinito numerable de las leyes de De Morgan (1.6).

9. Sea $\Omega = \mathbf{R} = (-\infty, \infty)$ y considere los siguientes conjuntos

$$A = (0, 3.5), \quad B = (2, 4], \quad C = [5, 8].$$

Obtenga $A \cup B \cup C$, $A \cap B \cap C$, $(A \cup B \cup C)^c$ y $(A \cap B \cap C)^c$.

10. Pruebe que cualquier tipo de intervalo en $\mathcal{A}_1, \ldots, \mathcal{A}_8$, se construye en términos de los intervalos en $\mathcal{A}_6$, bajo las operaciones básicas de conjuntos (unión, intersección y complemento); donde las familias de los intervalos están dadas en (1.9). Por lo tanto, la σ-álgebra de Borel satisface (1.8).

11. Considere el experimento de lanzar una moneda $n = 3$ veces donde se registra la cara superior al caer:

$$\Omega = \{xyz : x, y, z = a, s\} = \{aaa, aas, asa, ass, saa, sas, ssa, sss\}.$$

Véanse ejemplos 1.2.4 y 1.3.2.

a) Escriba explícitamente su σ-álgebra potencia; la cual se conforma de $2^{\#\Omega} = 2^8 = 256$ subconjuntos de Ω.

b) Escriba la probabilidad asignada a cada uno de dichos eventos, donde todos los eventos simples son equiprobables.

12. Sean A, B, C tres eventos de un espacio muestral Ω. Pruebe que

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

13. Sea $A_1, \ldots, A_n$; $n \geq 2$, una colección finita de eventos de un espacio muestral Ω. Demuestre

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j}^{n} P(A_i \cap A_j) + \sum_{i \neq j \neq k \neq i}^{n} P(A_i \cap A_j \cap A_k)$$
$$- \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n),$$

Sugerencia: aplique inducción matemática.

14. Sean A y B dos eventos tales que $A \subset B$, ¿Cuál es la probabilidad de los eventos $A \cup B$, $A \cap B$ y $A \cap B^c$?

15. Considere a la sucesión decreciente de eventos

$$A_1 \supset A_2 \supset \cdots \supset A,$$

con $A = \bigcap_{n=1}^{\infty} A_n$. Pruebe que

$$A_1^c \subset A_2^c \subset \cdots \subset A^c.$$

Recíprocamente, si

$$A_1 \subset A_2 \subset \cdots \subset A,$$

con $A = \bigcup_{n=1}^{\infty} A_n$, entonces

$$A_1^c \supset A_2^c \supset \cdots \supset A^c.$$

Sugerencia: aplique las leyes de De Morgan.

16. Pruebe que el axioma aditivo numerable (1.11) equivale a cualquiera de las siguientes propiedades de continuidad de la probabilidad.

- Continuidad creciente (1.23).
- Continuidad decreciente (1.24).
- Continuidad en Ω: si $A_n \uparrow \Omega$, $n \to \infty$, entonces $\lim_{n\to\infty} P(A_n) = 1$.
- Continuidad en el vacío: si $A_n \downarrow \varnothing$, $n \to \infty$, entonces $\lim_{n\to\infty} P(A_n) = 0$.

Por lo tanto, dichas propiedades de continuidad pueden sustituir al axioma aditivo numerable en la definición de espacio de probabilidad.

17. Considere el espacio de probabilidad continuo definido en el Ejemplo 1.3.4.4, donde

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

a) Compruebe que esta función satisface los requisitos (1.14):

$$f(x) \geq 0; \quad \text{para} \quad x \in \mathbf{R} \quad \text{y} \quad \int_{-\infty}^{\infty} f(t)\, \mathrm{d}t = 1.$$

b) Calcule las siguientes probabilidades: $P((-\infty, 0))$, $P((0, \infty))$, $P((-\infty, 3))$, $P((3, \infty))$, $P((-\infty, 3) \cup (3.5, \infty))$.

18. El formato de placa de automóvil sedán en el Estado de Aguascalientes es

A	u	v	$-$	w	x	$-$	y	z

donde u y v son cualesquiera de las 26 letras mayúsculas del alfabeto latino y w, x, y, z son dígitos del 0 al 9. Determine el total de placas disponibles para dicha entidad.

19. Sean los enteros no negativos n, n_1, n_2, k tales que $0 \leq k \leq n_1, n_2$ y $n_1+n_2 = n$. Demuestre que

$$\binom{n}{k} = \binom{n_1}{0}\binom{n_2}{k} + \binom{n_1}{1}\binom{n_2}{k-1} + \cdots + \binom{n_1}{k-1}\binom{n_2}{1} + \binom{n_1}{k}\binom{n_2}{0}.$$

20. Una caja contiene $1,000$ focos. La probabilidad de que haya al menos un foco defectuoso en la caja es de 0.1 mientras que la probabilidad de que haya al menos dos focos defectuosos es de 0.05. Encuentre la probabilidad de los siguientes eventos

    a) La caja no contiene focos defectuosos.

    b) La caja no contiene exactamente un foco defectuoso.

    c) La caja contiene a lo más un foco defectuoso.

21. Considere el experimento de observar la resistencia de $n = 7$ cilindros de ensayo de concreto. El conjunto de resultados posibles es

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_7) : \omega_i = F, E\},$$

donde E denota la falla de un cilindro durante la prueba.
(a) Verifique $\#\Omega = 128$.
(b) Obtenga los eventos simples de cada uno de los siguientes eventos

$$\begin{aligned}
&A_0 : 0 \text{ éxitos} = \text{no fallan los cilindros},\\
&A_1 : 1 \text{ éxito} = \text{falla un cilindro},\\
&A_2 : 2 \text{ éxitos} = \text{fallan dos cilindros},\\
&A_3 : 3 \text{ éxitos} = \text{fallan tres cilindros},\\
&A_4 : 4 \text{ éxitos} = \text{fallan cuatro cilindros},\\
&A_5 : 5 \text{ éxitos} = \text{fallan cinco cilindros},\\
&A_6 : 6 \text{ éxitos} = \text{fallan seis cilindros},\\
&A_7 : 7 \text{ éxitos} = \text{fallan todos los cilindros}.
\end{aligned}$$

(c) Verifique

$$\#A_k = \binom{7}{k}; \quad \text{para} \quad k = 0, 1, \ldots, 7.$$

(d) Verifique la igualdad

$$2^7 = \binom{7}{0} + \binom{7}{1} + \cdots + \binom{7}{7}.$$

(e) Asuma que, para cualquier cilindro particular, la probabilidad de falla es $p = P(E) = 0.004$. Obtenga la probabilidad $P(FFFFFFF)$, de que todos los cilindros hayan pasado la prueba de resistencia.
(f) Asuma que una obra se destruye y reconstruye si cuando menos hay dos cilindros de ensayo defectuosos. ¿Qué porcentaje de las obras tendrán que ser reconstruidas?

22. Proporcione un ejemplo de tres eventos A, B, C, que satisfaga (1.35), pero que no sean independientes.

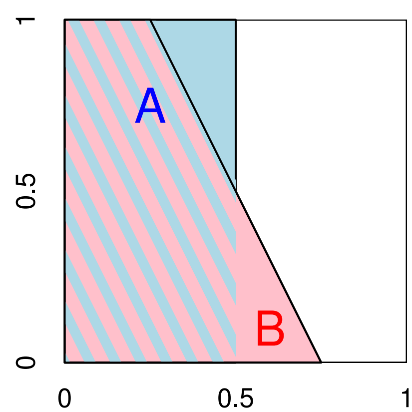
23. Considere el espacio de probabilidad uniforme (continuo) en el cuadrado unitario del plano

$$\Omega = \{(x, y) : 0 < x, y < 1\}.$$

   a) Verifique si los eventos A y B son independientes o no, con

$$A = \left\{(x, y) : 0 < x < \frac{1}{2},\ 0 < y < 1\right\},$$
$$B = \left\{(x, y) : 0 < x < \frac{3}{4},\ 0 < y < 1,\ 0 < 4x + 2y < 3\right\}$$

   En la siguiente figura se muestra la gráfica de ambos eventos.



   b) Obtenga las probabilidades incondicional y condicionales: $P(B)$, $P(B \mid A)$ y $P(B \mid A^c)$

   c) Si ocurre el evento A, ¿cuál es su predicción del evento B?

   d) Si ocurre el evento A^c, ¿cuál es su predicción del evento B?

24. Considere el experimento en dos etapas de la urna de Polya; véase el Ejemplo 1.7.3. En cada ensayo se extrae una bola y se registra su color. Para el primer ensayo, la urna contiene r bolas rojas y b negras. Para el segundo ensayo, se agregan $c > 0$ bolas del color de la bola de la primera extracción. Obtenga las probabilidades incondicional y condicionales: $P(R_1)$, $P(R_1 \mid R_2)$ y $P(R_1 \mid B_2)$, donde

R_1 : primera bola es roja, $\quad B_1$: primera bola es negra,

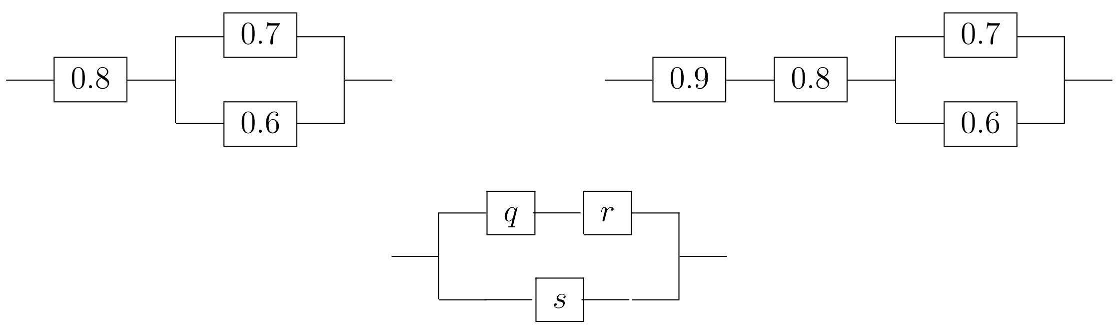
R_2 : segunda bola es roja, $\quad B_2$: segunda bola es negra.

25. a) Un circuito en *serie* de dos componentes funciona si y sólo sí, ambos componentes funcionan. Asuma que cada dispositivo funciona independiente uno del otro, con probabilidad p especificada en la siguiente figura. Calcule la probabilidad de que falle el circuito en serie.

b) Un circuito en *paralelo* de dos componentes funciona si al menos uno de ellos funciona, como se describe en la siguiente figura. Cada dispositivo falla independiente del otro con igual probabilidad q, ¿cuál es el valor de q de modo que este circuito tenga la misma confiabilidad que la del circuito en serie del inciso anterior?

serie paralelo

26. Encuentre la confiabilidad de los circuitos mixtos descritos en la siguiente figura. Cada componente funciona independiente de los otros, con probabilidad especificada en dicha figura.



Capítulo 2

Variables aleatorias discretas

En este capítulo se definirá el concepto de variable aleatoria. Se analizarán las propiedades básicas de las variables aleatorias discretas.

2.1. Variable aleatoria

En esta sección se da una introducción a las variables aleatorias. Por el contexto del capítulo, se abordan principalmente ejemplos de variables aleatorias de naturaleza discreta. En el siguiente capítulo se describe a las variables aleatorias del tipo continuo.

El concepto de *variable aleatoria* simplifica la descripción de eventos de interés de un espacio muestral Ω, derivado de un experimento. Por ejemplo, el clima de una región, en una hora y fecha precisa, se determina por una amplia gama de factores o variables, como la temperatura, presión atmosférica, a nivel del suelo y a otras alturas superiores, presencia y tipo de nubes, humedad relativa, velocidad y dirección del viento, intensidad de lluvia, entre otros. Este es un espacio muestral complejo, que se puede simplificar en términos del modelo de probabilidad variable aleatoria.

La primera característica a identificar de una variable aleatoria de interés X, es su *conjunto de valores posibles*. Ello determina su naturaleza: continua, discreta u otro tipo. En el caso de la temperatura X que habrá el día de mañana a las 08:00, se considera que toma cualquier valor en un amplio intervalo, como $-30^\circ < X < 50^\circ$ (grados centígrados). Incluso se puede asumir que toma cualquier valor de recta real: $-\infty < X < \infty$. Informalmente, si una variable aleatoria toma cualquier valor de un intervalo, se dice que es de naturaleza *continua*. En cambio, si el conjunto de valores posibles es finito o infinito numerable, como $X = 0, 1, \ldots,$ entonces se dice que X es una variable aleatoria del tipo o naturaleza *discreta*. Los modelos de variables aleatorias continuas o discretas explican gran parte de los fenómenos aleatorios de interés.

La *distribución* de probabilidad de una variable aleatoria X determina la probabilidad de sus eventos asociados. Para el ejemplo de la temperatura, es de interés la probabilidad del evento helada: $\{X \leq 0\}$. También, la distribución establece una predicción puntual. Por ejemplo, el valor de la temperatura será su *media poblacional* o *valor esperado* $\mu = \mathrm{E}\, X = 18.34^\circ$. El siguiente ejemplo describe un modelo de variable aleatoria discreta.

Ejemplo 2.1.1 (juego de la ruleta)**.** Considere a un jugador de la ruleta que apuesta $n = 5$ veces a los números rojos. En cada ensayo de Bernoulli, hay *éxito* (rojo) o *fracaso* (no rojo), donde la probabilidad de éxito es

$$p = P(E) = \frac{18}{38} = 0.4737.$$

El espacio muestral Ω tiene $2^n = 2^5 = 32$ resultados posibles:

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) : \omega_i = \text{fracaso ó éxito}\}\\ &= \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) : \omega_i = F, E\}\\ &= \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) : \omega_i = 0, 1\}.\end{aligned}$$

Tanto para el jugador como la casa, los resultados posibles del experimento se resumen con los valores posibles de la *variable aleatoria* X:

$$\begin{aligned}X(\omega) &: \text{número de éxitos en 5 ensayos}\\ &= \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5, \quad \text{para cada} \quad \omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5) \in \Omega.\end{aligned}$$

En la Tabla 2.1 se muestra el valor de esta variable según cada evento simple. El *conjunto de valores posibles* de la variable aleatoria X es $\{0, 1, 2, 3, 4, 5\}$. Se denota también como

$$X = 0, 1, 2, 3, 4, 5.$$

Al establecer el conjunto de resultados posibles, se descartan otros valores numéricos, como $X = -2, 3.5, 6$. Note también que el modelo de variable aleatoria es más práctico que el de espacio muestral, pues el número de resultados posibles de X es mucho menor que el número de eventos simples de Ω. Luego, por ser de interés, se calcula la probabilidad de cada valor posible de la variable aleatoria X. En este sentido, note que $\{X = 0\} = \{FFFFF\}$. Este evento simple es la intersección de cinco eventos independientes. Aquí se asume la independencia de los resultados de los ensayos de Bernoulli. Por (1.39), se tiene

$$\begin{aligned}P(X = 0) &= P(FFFFF) = P(F)P(F)P(F)P(F)P(F)\\ &= [P(F)]^5 = (1 - p)^5\\ &= (1 - 0.4737)^5 = 0.5263^5\\ &= 0.0404.\end{aligned}$$

Así mismo, el evento $\{X = 1\}$ es la unión de 5 eventos simples excluyentes y equiprobables; véase la Tabla 2.1. Por lo que

$$\begin{aligned}P(X = 1) &= P(EFFFF, FEFFF, FFEFF, FFFEF, FFFFE)\\ &= P(EFFFF) + P(FEFFF) + P(FFEFF) + P(FFFEF) + P(FFFFE)\\ &= 5P(EFFFF) = 5P(E)[P(F)]^4\\ &= \binom{5}{1} p^1 (1 - p)^4 = 5(0.4737^1)0.5263^4\\ &= 0.1817.\end{aligned}$$

El coeficiente binomial $\binom{5}{1} = 5$, explica el número de maneras de ubicar la letra E en las cinco posiciones disponibles. Por otro lado, para el evento $\{X = 2\}$, se requiere un arreglo de 5 ensayos con dos letras E. Sus dos posiciones se seleccionan de $\binom{5}{2} = 10$ diferentes maneras. Entonces

$$P(X = 2) = \binom{5}{2} P(EEFFF) = 10p^2(1-p)^3 = 10 \times 0.4737^2 \times 0.5263^3 = 0.3271.$$

Análogamente

$$P(X = 3) = \binom{5}{3} P(EEEFF) = 10p^3(1-p)^2 = 10 \times 0.4737^3 \times 0.5263^2 = 0.2944,$$
$$P(X = 4) = \binom{5}{4} P(EEEEF) = 5p^4(1-p) = 5 \times 0.4737^4 \times 0.5263 = 0.1325,$$
$$P(X = 5) = \binom{5}{5} P(EEEEE) = p^5 = 0.4737^5 = 0.0238.$$

La *función de densidad* resume estas probabilidades

$$f(x) = P(X = x) = \binom{5}{x} p^x (1-p)^{5-x} = \binom{5}{x} 0.4737^x 0.5263^{5-x} = \begin{cases} 0.0404 & x = 0 \\ 0.1817 & x = 1 \\ 0.3271 & x = 2 \\ 0.2944 & x = 3 \\ 0.1325 & x = 4 \\ 0.0238 & x = 5. \end{cases} \tag{2.1}$$

La respectiva *función de distribución* aporta las probabilidades acumuladas:

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 0.0404 & 0 \leq x < 1 \\ 0.0404 + 0.1817 = 0.2221 & 1 \leq x < 2 \\ 0.2221 + 0.3271 = 0.5493 & 2 \leq x < 3 \\ 0.5493 + 0.2944 = 0.8437 & 3 \leq x < 4 \\ 0.8437 + 0.1325 = 0.9762 & 4 \leq x < 5 \\ 0.9762 + 0.0238 = 1 & x \geq 5 \end{cases} \tag{2.2}$$

La Figura 2.1 muestra la gráfica de las funciones de densidad $f(x)$ y de distribución $F(x)$ de la variable aleatoria X. Ambas funciones determinan su *distribución*. Por ejemplo, se deduce lo siguiente.

1. El conjunto de valores posibles de la variable aleatoria X son los puntos x de la recta real donde la función de densidad $f(x)$ sea positiva: $x = 0, 1, 2, 3, 4, 5$. Se descarta cualquier otro punto como valor posible.

2. El conjunto de valores posibles de la variable aleatoria X son los puntos de la recta real donde la función de distribución $F(x)$ tiene un salto: $x = 0, 1, 2, 3, 4, 5$. Se descarta cualquier otro valor posible. La magnitud del salto en x es la probabilidad $f(x) = P(X = x)$.

3. Los valores posibles de la variable aleatoria X no son equiprobables. De hecho, es más factible el evento $\{X = 0\}$ que $\{X = 5\}$. Una comparación similar aplica entre los eventos $\{X = 1\}$ y $\{X = 4\}$, así como entre $\{X = 2\}$ y $\{X = 3\}$. Lo que evidencia la ventaja de la casa sobre el jugador.

4. Con ambas funciones se calcula la probabilidad de cualquier de evento de interés. Por ejemplo, se obtiene el intervalo de predicción $1 \leq X \leq 4$, con cobertura de 93.58 %:

$$\begin{aligned} P(1 \leq X \leq 4) &= P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) \\ &= f(1) + f(2) + f(3) + f(4) \\ &= 0.1817 + 0.3271 + 0.2944 + 0.1325 \\ &= 0.9357 = 93.57\,\%. \end{aligned}$$

   y

$$\begin{aligned} P(1 \leq X \leq 4) = P(0 < X \leq 4) &= P(X \leq 4) - P(X \leq 0) \\ &= F(4) - F(0) = 0.9762 - 0.0404 \\ &= 0.9358 = 93.58\,\%. \end{aligned}$$

5. Otros eventos de interés son ganar o perder. El último evento es más factible que el primero:

$$P(\text{perder en sesión de 5 juegos}) = P(X \leq 2) = F(2) = 0.5493 = 54.93\,\%,$$

   y

$$\begin{aligned} P(\text{ganar en sesión de 5 juegos}) &= P(X \geq 3) = 1 - P(X < 3) \\ &= 1 - P(X \leq 2) = 1 - F(2) \\ &= 1 - 0.5493 = 0.4507. \end{aligned}$$

   Por último, note que hay un mayor riesgo de perder con la estrategia de jugar $n = 5$ veces que sólo una vez. De hecho, compare la probabilidad de ganar en ambos escenarios: $0.4507 < 0.4737$.

distribución discreta!binomial
distribución discreta!binomial
El adjetivo “aleatorio” de una variable se debe a lo incierto de su valor numérico, transmitido de la incertidumbre del resultado del experimento. En la Tabla 2.2 se muestran algunos ejemplos de variables aleatorias de la vida cotidiana, así como su tipo.

Según la relación funcional $\omega \to X(\omega)$, si el resultado del experimento es el evento simple $\omega \in \Omega$, entonces el valor de la variable aleatoria es $X(\omega) \in \mathbf{R}$. Usualmente se ignora

ω	$X(\omega)$	ω	$X(\omega)$	ω	$X(\omega)$	ω	$X(\omega)$
00000	0	10010	2	11100	3	01011	3
10000	1	10001	2	11010	3	00111	3
01000	1	01100	2	11001	3	11110	4
00100	1	01010	2	10110	3	11101	4
00010	1	01001	2	10101	3	11011	4
00001	1	00110	2	10011	3	10111	4
11000	2	00101	2	01110	3	01111	4
10100	2	00011	2	01101	3	11111	5

Tabla 2.1. Variable aleatoria suma de éxitos en $n = 5$ ensayos de Bernoulli.

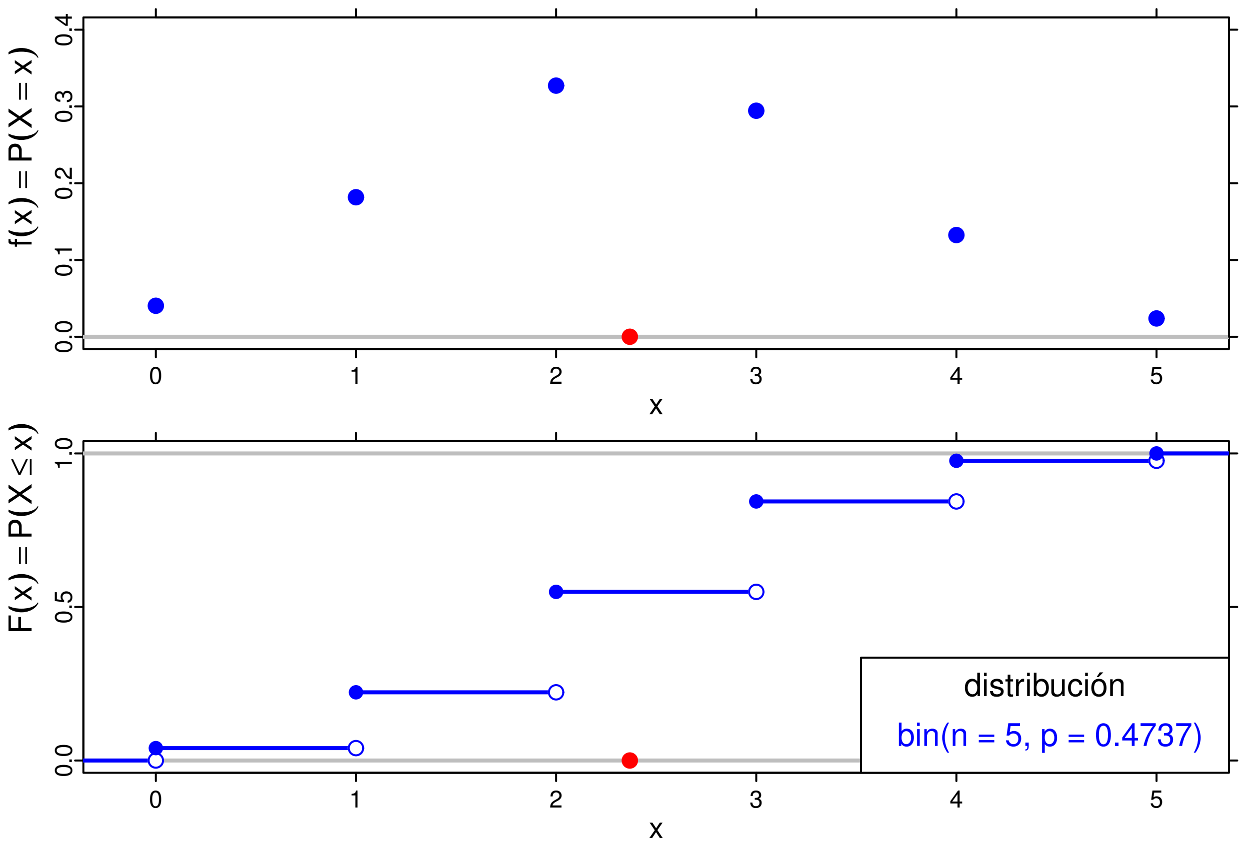


Figura 2.1. Gráficas de las funciones de densidad y de distribución binomial; Ejemplo 2.1.1.

variable aleatoria X	valores posibles	tipo
número varones de una familia de cinco hijos	$0,1,2,3,4,5$	discreta
número de juegos en la ruleta hasta ganar por primera vez	$1,2,\ldots$	discreta
número de fracasos en la ruleta hasta ganar por primera vez	$0,1,\ldots$	discreta
número de clientes que llegan a un banco antes de abrir	$0,1,\ldots$	discreta
duración en horas enteras de carga de batería de un teléfono	$0,1,\ldots$	discreta
duración en horas de la carga de dicha batería	$(0,\infty)$	continua
temperatura para el día de mañana a las 08:00 horas	$(-20,30)^\circ$ C	continua
pérdida de volumen de aceite de motor por uso en tres meses	$(0,100)\,\%$	continua
estatura de un estudiante universitario seleccionado al azar	$(0,\infty)$	continua
error de un tornillo de ancho 5 cm, elegido al azar de un lote	$(-\infty,\infty)$	continua

Tabla 2.2. Ejemplos de variables aleatorias, su conjunto de valores posibles y tipo.

el resultado del experimento ω, así como dicha regla de asociación. Por el contrario, en la práctica se dispone sólo del valor $X(\omega)$. Por ejemplo, si X denota a la temperatura del día de mañana a las 08:00 horas en un lugar específico, el valor de esta variable dependerá de todos los indicadores o factores que influyen en el clima de mañana: presión barométrica, dirección y velocidad de los vientos, humedad relativa, presencia o ausencia de nubes, radiación ultravioleta, día del año, altura, la temperatura misma en las horas de la madrugada, entre otros factores no observables. Se registrará la temperatura $X(\omega)$, a pesa de ignorar el correspondiente clima $\omega \in \Omega$, junto con la mencionada relación funcional $\omega \to X(\omega)$. Cabe aclarar que la tendencia o predicción de la temperatura para las 08:00 horas de mañana se apoya de sofisticados instrumentos de medición meteorológica, así como de modelos físico-meteorológicos-estocásticos.

Por todo lo anterior, se establece un espacio de probabilidad paralelo al original $(\Omega,\mathcal{A},P)$, que modela específicamente la incertidumbre de la variable aleatoria X. Con este fin, el nuevo espacio de probabilidad se construye en la recta real: $(\mathbf{R},\mathcal{B}(\mathbf{R}),P_X)$, donde P_X denota la *distribución* de probabilidad de la variable aleatoria X. Esta distribución se determina por la *función de distribución* $F(x)=P(X\leq x)$; $x\in\mathbf{R}$. Por ejemplo, recuerde la función (2.2) del Ejemplo 2.1.1.

Ahora se definirá de manera formal el concepto de variable aleatoria y su función de distribución.

Definición 2.1.2. Sean $(\Omega,\mathcal{A},P)$ un espacio de probabilidad y $X:\Omega\to\mathbf{R}$, una función real definida en el espacio muestral Ω. Se dice que X es una *variable aleatoria* si

$$\{X\leq x\}\doteq\{\omega\in\Omega: X(\omega)\leq x\}\in\mathcal{A},\quad \text{para todo}\quad x\in\mathbf{R}. \tag{2.3}$$

Su *función de distribución* es

$$F(x)\doteq F_X(x)\doteq P(X\leq x);\quad x\in\mathbf{R}. \tag{2.4}$$

El argumento de letra minúscula invoca la variable aleatoria de letra mayúscula respectiva. Para evitar confusión, ante la presencia de dos o más variables aleatorias, se recomienda el uso de subíndice en la función de distribución.

Antes de llevarse a cabo un experimento, hay incertidumbre o ignorancia del valor que tomará la variable aleatoria X. Su distribución permite el cálculo de la probabilidad de eventos de interés asociados a esta variable aleatoria. Lo que prevé o descarta posibles resultados. Para el ejemplo de la temperatura, puede resultar de interés la probabilidad de los eventos $\{X \leq 0\}$, $\{X > 0\}$, $\{10 < X \leq 22\}$, $\{X > 35\}$. En particular, el evento $\{X \leq 0\}$ significa helada. Este clima afectará los cultivos de hortalizas de una zona agrícola. Si la probabilidad de este evento es muy alta, entonces hay que tomar medidas precautorias para proteger los cultivos. En el contexto de seguros de siniestros agrícolas por causa de heladas, la probabilidad del evento $\{X \leq 0\}$ mide el riesgo de la cobertura del daño. A mayor probabilidad de tal evento, mayor será el riesgo, así como el valor de la prima. Otro evento de interés relacionado con la temperatura es $\{X \leq 5\}$, que se entiende como un detonador de los contagios de gripa. Por lo cual, el cálculo de su probabilidad ayuda a prevenir a los padres de familia o a las autoridades sanitarias del riesgo de brotes de epidemias de gripa. Así mismo, en días calurosos es de interés calcular la probabilidad de eventos como $\{X > 35\}$, para advertir casos de deshidratación en menores y adultos mayores.

Note que en (4.7) se invocan intervalos de la forma $(-\infty, x]$; para $x \in \mathbf{R}$. Por (1.7)-(1.9) del Ejemplo 1.2.8, esta familia de intervalos genera a la σ-álgebra de Borel $\mathcal{B}(\mathbf{R})$, la cual contiene toda clase de intervalos. Por lo tanto, con la función de distribución (2.4) se calcula la probabilidad de cualquier evento de interés de la variable aleatoria X en términos de intervalos; véase la Tabla 2.3. Se describirán algunos ejemplos de eventos de la variable aleatoria X en términos de eventos de la forma (4.7). Para $a, b \in \mathbf{R}$, con $a \leq b$, se tiene

$$\begin{aligned}
\{X > b\} &= \{X \leq b\}^c, \\
\{X < b\} &= \bigcup_{n\geq 1} \left\{X \leq b - \frac{1}{n}\right\}, \\
\{a < X \leq b\} &= \{X \leq b\} \cap \{X > a\} = \{X \leq b\} \cap \{X \leq a\}^c, \\
\{X = b\} &= \bigcap_{n\geq 1} \left\{b - \frac{1}{n} < X \leq b\right\} = \{X \leq b\} \cap \bigcap_{n\geq 1} \left\{X \leq b - \frac{1}{n}\right\}^c.
\end{aligned}$$

Así, la probabilidad de los eventos anteriores queda en términos de la función de distribución $F(x)$. Para los primeros dos eventos, se tiene

$$\begin{aligned}
&P(X > b) = P(\{X \leq b\}^c) = 1 - P(X \leq b) = 1 - F(b), \\
&P(X < b) = P\left(\bigcup_{n\geq 1} \left\{X \leq b - \frac{1}{n}\right\}\right) = \lim_{n\to\infty} P\left(X \leq b - \frac{1}{n}\right) = \lim_{n\to\infty} F\left(b - \frac{1}{n}\right) = F(b-).
\end{aligned}$$

La segunda y última igualdad de la expresión anterior se debe a la propiedad de continuidad de la probabilidad (1.23), que se retoma en (2.7), más adelante. Por otro lado, una importante identidad es

$$P(a < X \leq b) = F(b) - F(a); \quad \text{para} \quad -\infty < a \leq b < \infty. \tag{2.5}$$

$$\begin{array}{llll} \{X=a\}, & \{X>a\}, & \{a\le X<b\}, & \\ \{X\le a\}, & \{X\ge a\}, & \{a<X<b\}, & \\ \{X<a\}, & \{a\le X\le b\}, & \{a<X\le b\}, & \text{con} \quad -\infty<a\le b<\infty. \end{array}$$

Tabla 2.3. Algunos eventos de interés asociados a una variable aleatoria X.

Esta fórmula aplica incluso para los casos $a=-\infty$ o $b=\infty$; con la aclaración de que $P(X=\pm\infty)=0$ pues, por definición, una variable aleatoria es una función real. Para verificar tal identidad, considere la partición del evento $\{X\le b\}$ en dos eventos excluyentes:

$$\{X\le b\}=\{X\le a\}\cup\{a<X\le b\}.$$

Esta partición es análoga a la del semieje cerrado $(-\infty,b]$:

$$(-\infty,b]=(-\infty,a]\cup(a,b].$$

Así, se obtiene (2.5):

$$P(X\le b)=P(X\le a)+P(a<X\le b),$$
$$P(a<X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a).$$

Por último, al aplicar la propiedad de continuidad de la probabilidad (1.24) y la identidad (2.5), se tiene

$$\begin{aligned} P(X=b) &= P\left(\bigcap_{n\ge 1}\left\{b-\frac{1}{n}<X\le b\right\}\right)=\lim_{n\to\infty}P\left(b-\frac{1}{n}<X\le b\right) \\ &= \lim_{n\to\infty}\left[F(b)-F\left(b-\frac{1}{n}\right)\right]=F(b)-F(b-); \quad \text{para} \quad b\in\mathbf{R}. \end{aligned}$$

Se concluye

$$P(X=x)=F(x)-F(x-); \quad \text{para} \quad x\in\mathbf{R}. \tag{2.6}$$

Esta fórmula es relevante en el caso de una variable aleatoria *discreta*, en donde el salto de $F(x)$ define a la *función de densidad* $f(x)$.

Toda variable aleatoria tiene una única función de distribución, la cual satisface las tres propiedades del siguiente teorema. Recíprocamente, dada una función $F(x)$ que satisfaga dichas propiedades, entonces esta es la función de distribución de cierta variable aleatoria. Si dos variables aleatorias X y Y comparten su función de distribución, se dice que son *iguales en distribución*. Se denota como $X\overset{d}{=}Y$. Por ejemplo, si X y Y denotan el número de clientes que llegan a una gasolinera en dos semanas consecutivas, entonces, ambas variables aleatorias pueden ser iguales en distribución.

Teorema 2.1.3. Sea $X:\Omega\to\mathbf{R}$ una variable aleatoria con función de distribución $F(x)=P(X\le x)$; $x\in\mathbf{R}$. Entonces

1. $F(x)$ es creciente (en el sentido amplio)

$$F(x)\le F(y); \quad x\le y.$$

2. $F(x)$ es continua por la derecha

$$F(x) = F(x+) \circeq \lim_{y \to x+} F(y). \tag{2.7}$$

3.
$$F(-\infty) \circeq \lim_{x \to -\infty} F(x) = 0 \quad \text{y} \quad F(\infty) \circeq \lim_{x \to \infty} F(x) = 1. \tag{2.8}$$

Recíprocamente, si $F(x)$ es una función real que satisface tales tres propiedades, entonces $F(x)$ corresponde a la función de distribución de una variable aleatoria X.

Demostración. El recíproco se verifica en el Teorema 3.7.1, más adelante.
1. Como
$$\{X \leq x\} \subset \{X \leq y\}; \quad \text{para} \quad x \leq y,$$
entonces
$$F(x) = P(X \leq x) \leq P(X \leq y) = F(y).$$
2. La igualdad (2.7), se obtiene con la propiedad de continuidad de la probabilidad (1.24). Para $x \in \mathbf{R}$, sea $\{x_n\}_{n \geq 1}$ una sucesión de números reales que decrece a x:

$$x_1 \geq x_2 \geq \cdots \geq x_n \geq \cdots > x \quad \text{y} \quad \lim_{n \leftrightarrow \infty} x_n = x.$$

Se denota como $x_n \downarrow x$, cuando $n \to \infty$. Entonces, la sucesión de eventos $\{X \leq x_n\}$, decrece al evento intersección $\{X \leq x\}$:

$$\{X \leq x_1\} \supset \{X \leq x_2\} \supset \cdots \supset \bigcap_{n \geq 1} \{X \leq x_n\} = \{X \leq x\}.$$

Por la propiedad de continuidad de la probabilidad (1.24), se tiene

$$\lim_{n \to \infty} F(x_n) = \lim_{n \to \infty} P(X \leq x_n) = P(X \leq x) = F(x).$$

Como la sucesión $\{x_n\}_{n \geq 1}$ es arbitraria, se obtiene la continuidad por la derecha de $F(x)$:

$$F(x+) = \lim_{y \to x+} F(y) = F(x).$$

3. Sean
$$B_n = \{X \leq -n\} \quad \text{y} \quad D_n = \{X \leq n\}; \quad n \geq 1.$$
Similar al inciso anterior, $\{B_n\}_{n \geq 1}$ es una sucesión de eventos que decrece al conjunto vacío:

$$\bigcap_{n \geq 1} B_n = \bigcap_{n \geq 1} \{X \leq -n\} = \varnothing.$$

Por (1.24), se tiene

$$F(-\infty) = \lim_{n \to \infty} F(-n) = \lim_{n \to \infty} P(X \leq -n) = \lim_{n \to \infty} P(B_n) = P(\varnothing) = 0.$$

En cambio, $\{D_n\}_{n\geq 1}$ es una sucesión que crece al espacio muestral Ω:

$$\bigcup_{n\geq 1} D_n = \bigcup_{n\geq 1}\{X \leq n\} = \{X < \infty\} = \Omega.$$

Por la propiedad de continuidad de la probabilidad (1.23), resulta

$$F(\infty) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} P(X \leq n) = \lim_{n\to\infty} P(D_n) = P(\Omega) = 1.$$

□

Por otro lado, al considerar (2.7), la función de distribución $F(x)$ es continua en $x \in \mathbf{R}$ si y solo si es continua por la izquierda en dicho punto. En cualquier caso, la función $F(x)$ tiene límite por la izquierda, como lo afirma la siguiente proposición. La propiedad de continuidad de la probabilidad (1.23) tendrá un rol relevante en la demostración.

Proposición 2.1.4. Sea $X : \Omega \to \mathbf{R}$ una variable aleatoria con función de distribución $F(x) = P(X \leq x)$; $x \in \mathbf{R}$. Entonces $F(x)$ tiene límite por la izquierda y

$$F(x-) = P(X < x) \leq P(X \leq x) = F(x); \quad \text{para} \quad x \in \mathbf{R}. \tag{2.9}$$

Demostración. Es suficiente verificar la primera igualdad en (2.9). Considere $\{x_n\}_{n\geq 1}$, una sucesión creciente de números reales que converge a x, $x_n \uparrow x$, cuando $n \to \infty$:

$$x_1 \leq x_2 \leq \cdots \leq x_n \leq \cdots < x \quad \text{y} \quad \lim_{n\leftrightarrow\infty} x_n = x.$$

La sucesión de eventos $\{X \leq x_n\}_{n\geq 1}$ es creciente y converge a $\{X < x\}$:

$$\{X \leq x_1\} \subset \{X \leq x_2\} \subset \cdots \subset \bigcup_{n\geq 1}\{X \leq x_n\} = \{X < x\}.$$

Por la propiedad de continuidad de la probabilidad (1.23), se tiene

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(X \leq x_n) = P(X < x).$$

Como es arbitraria la sucesión $\{x_n\}_{n\geq 1}$, entonces el límite anterior es $F(x-)$. Por lo tanto, $F(x)$ tiene límite por la izquierda de x y satisface la primera igualdad en (2.9):

$$F(x-) \mathrel{\hat{=}} \lim_{y\to x-} F(y) = P(X < x).$$

□

Gran parte de los fenómenos aleatorios se modelan con dos tipos de variables aleatorias: discreto o continuo.

- El primer tipo es el de una variable aleatoria *discreta*, donde su conjunto de valores posibles es finito o infinito numerable de $\mathbf{R}$. En este caso, la función de distribución $F(x)$ es discontinua por la izquierda sólo en dicho conjunto. Por ejemplo, recuerde la forma discontinua de la función de distribución (2.2), cuya gráfica se muestra en la Figura 2.1[abajo]. Otro

distribución	tipo	función de distribución
Poisson ($\mu = 2.5$)	discreta	$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ e^{-2.5} \sum_{i=0}^{\lfloor x \rfloor} \frac{2.5^i}{i!} & x \geq 0 \end{cases}$
normal estándar	continua	$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, \mathrm{d}\, u; \quad z \in \mathbf{R}$

Tabla 2.4. Dos ejemplos de distribución de probabilidad: Poisson (media $\mu = 2.5$) y normal estándar.

ejemplo se refiere a una variable aleatoria de distribución de *Poisson* con *media* $\mu = 2.5$. Su conjunto de valores posibles son los enteros no negativos. La función de distribución $F(x)$ se muestra en la Tabla 2.4 mientras que en la Figura 2.2[arriba] aparece su gráfica. Aquí se aprecian las características del Teorema 2.1.3: $F(x)$ es creciente, continua por la derecha, así como $F(-\infty) = 0$ y $F(\infty) = 1$. Además, $F(x)$ es una función continua, excepto en sus valores posibles. De hecho, como lo afirma la Proposición 2.1.4, $F(x)$ tiene límite por la izquierda.

- El segundo tipo se refiere a una variable aleatoria *continua*, donde su función de distribución $F(x)$ es continua en todo punto x de la recta real. Informalmente, una variable aleatoria es continua si puede tomar cualquier valor en un intervalo. Recuerde el ejemplo de la temperatura al día de mañana a las 08:00 horas. Otro ejemplo corresponde a una variable aleatoria Z de distribución *normal estándar*, donde cualquier punto de la recta real z es un valor posible. La función de distribución $\Phi(z) = P(Z \leq z)$; para $z \in \mathbf{R}$, es continua, estrictamente creciente, con $\Phi(-\infty) = 0$ y $\Phi(\infty) = 1$; véase la Tabla 2.4. Su gráfica aparece en la Figura 2.2[abajo].

Definición 2.1.5. Una variable aleatoria X es *discreta* si su imagen es un conjunto finito o infinito numerable. Su *conjunto de valores posibles* es la colección de puntos de la recta real

$$\mathcal{X} = \{x \in \mathbf{R} : P(X = x) > 0\},$$

con $1 \leq \#\mathcal{X} \leq \#\mathbf{N}$. La *función de densidad* es

$$f(x) = P(X = x); \quad x \in \mathcal{X}. \tag{2.10}$$

En otros casos, la función de densidad es implícitamente cero: $f(x) = 0$; para $x \notin \mathcal{X}$.

Usualmente, el conjunto de valores posibles de una variable aleatoria discreta son los enteros no negativos o un subconjunto del mismo. En tal caso, se denota como

$$X = 0, 1, 2, \ldots$$

Un ejemplo de aplicación de variable aleatoria discreta es el número X de vehículos que pasan por un punto específico, durante un período de tiempo.

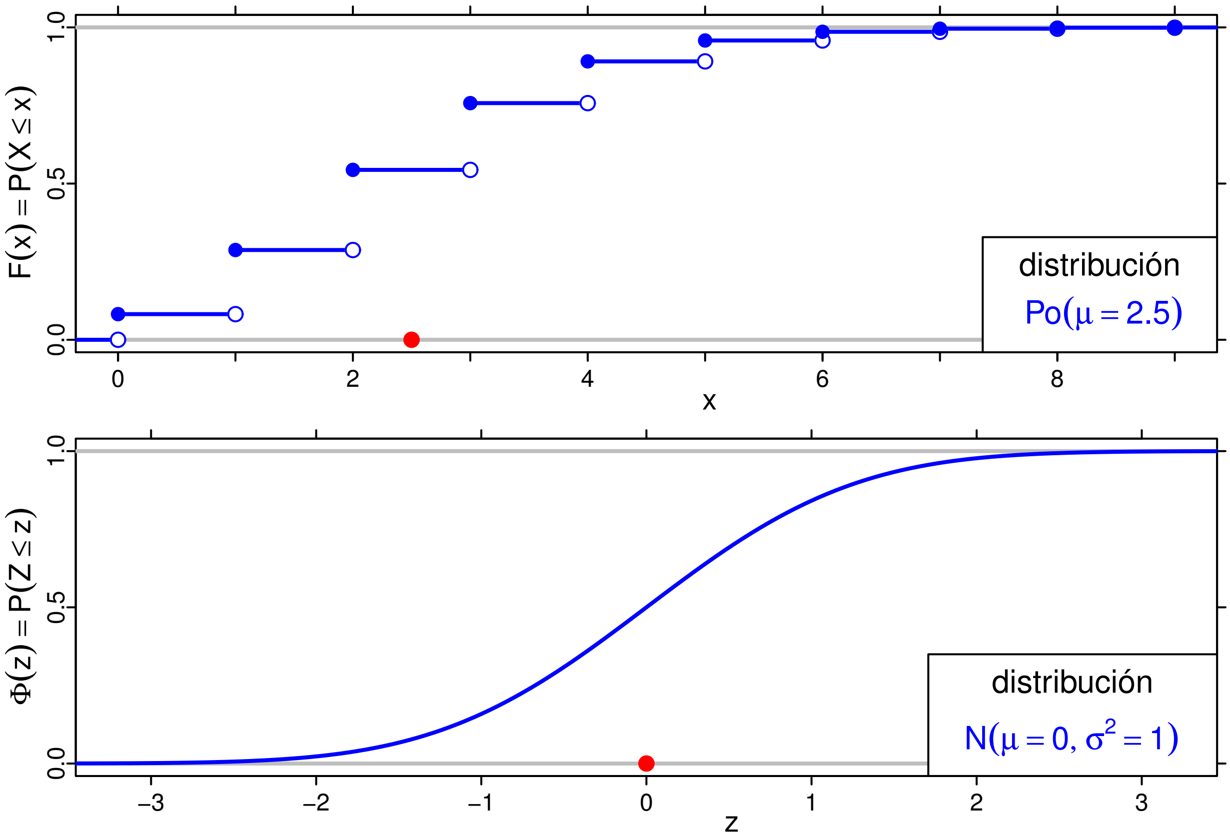


Figura 2.2. Gráficas de funciones de distribución de Poisson y normal estándar.

La función de densidad $f(x)$ de una variable aleatoria discreta X está determinada por su función de distribución $F(x)$, y viceversa. De hecho, por (2.6), se tiene

$$f(x) = P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(x-); \quad x \in \mathbf{R}. \tag{2.11}$$

Cabe recordar que la segunda igualdad se debe a (2.9) y la unión de eventos excluyentes:

$$\{X \leq x\} = \{X < x\} \cup \{X = x\}$$

y

$$F(x) = P(X \leq x) = P(X < x) + P(X = x) = F(x-) - f(x).$$

Recíprocamente, note que

$$\{X \leq x\} = \bigcup_{y \leq x} \{X = y\}; \quad x \in \mathbf{R}.$$

Esta es una unión numerable de eventos excluyentes. Aquí se asume implícitamente que $y \in \mathcal{X}$ pues, en caso contrario, se tendría $P(Y = y) = 0$. Así

$$F(x) = P(X \leq x) = P\left(\bigcup_{y \leq x} \{X = y\}\right) = \sum_{y \leq x} P(X = y) = \sum_{y \leq x} f_X(y); \quad x \in \mathbf{R}.$$

Como consecuencia, la distribución de una variable aleatoria discreta X se determina por su función de distribución o de densidad. En particular, su conjunto de valores posibles se caracteriza de dos maneras equivalentes:

1. Los puntos $x \in \mathbf{R}$ donde la función de densidad $f(x)$ sea positiva. En particular, si el conjunto de valores posibles es $X = 0, 1, 2, \ldots$, se tiene

$$\begin{aligned} f(x) &= P(X = x) = P(X \leq x) - P(X < x) \\ &= F(x) - F(x-) = F(x) - F(x-1); \quad \text{para} \quad x = 0, 1, 2, \ldots \end{aligned}$$

2. Los puntos $x \in \mathbf{R}$, donde la función de distribución $F(x)$ tenga un salto. En particular, si $X = 0, 1, 2, \ldots$, se tiene

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= P(X \in \{0, 1, \ldots, x\}) \\ &= P(X = 0) + P(X = 1) + \cdots + P(X = x) \\ &= f(0) + f(1) + \cdots + f(x); \quad \text{para} \quad x = 0, 1, 2, \ldots \end{aligned} \tag{2.12}$$

El siguiente teorema caracteriza con dos propiedades a la función de densidad de una variable aleatoria discreta.

Teorema 2.1.6. Sea X una variable aleatoria discreta, con función de densidad (2.10): $f(x) = P(X = x)$; para $x = 0, 1, \ldots$ Entonces se satisface

1.
$$f(x) \geq 0; \quad \text{para} \quad x = 0, 1, \ldots \tag{2.13}$$

2.
$$\sum_{x=0}^{\infty} f(x) = 1. \tag{2.14}$$

Recíprocamente, si $f(x)$ es una función real que cumpla ambas propiedades, entonces esta corresponde a la función densidad de cierta variable aleatoria discreta X, con conjunto de valores posibles $X = 0, 1, \ldots$

Demostración. ($\Rightarrow$). La propiedad (2.13) se deduce del hecho que $f(x)$ es una probabilidad. Por otro lado, (2.14) se obtiene al aplicar el tercer axioma de Kolmogorov (1.11):

$$1 = P(\Omega) = P(X < \infty) = P\left(\bigcup_{x=0}^{\infty} \{X = x\}\right) = \sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} f(x).$$

($\Leftarrow$). Recíprocamente, si una función real $f(x)$ satisface (2.13)-(2.14), defina la función $F(x)$ como el lado derecho de (2.12). Esta función satisface las tres propiedades del Teorema 2.1.3. Por lo tanto, $F(x)$ corresponde a la función de distribución de cierta variable aleatoria X, la cual es discreta; pues toma valores en los enteros no negativos. $\square$

Una constante $c \in \mathbf{R}$, es una variable aleatoria discreta. Defina $X(\omega) = c$; para $\omega \in \Omega$. En la siguiente tabla se verifica formalmente el estatus de X como variable aleatoria.

$x \in \mathbf{R}$	$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$	$F(x) = P(X \leq x)$
$x < c$	$\varnothing$	0
$x \geq c$	$\{X = c\} = \Omega$	1

La función de distribución tiene un único salto, en $x = c$:

$$F(x) = P(X \leq x) = \begin{cases} 1 & x \geq c \\ 0 & x < c. \end{cases}$$

La correspondiente función de densidad concentra toda la probabilidad en el punto $x = c$:

$$f(x) = P(X = x) = \begin{cases} 1 & x = c \\ 0 & x \neq c. \end{cases}$$

Las gráficas de las funciones de distribución y densidad se muestran en la Figura 2.3; para el caso $c = 2.5$.

2.2. Distribuciones de probabilidad discretas

En esta sección de describen algunos modelos paramétricos de variables aleatorias discretas. Se abordan los modelos clásicos de distribución Bernoulli, binomial, geométrica, Poisson, uniforme discreta y multinomial.

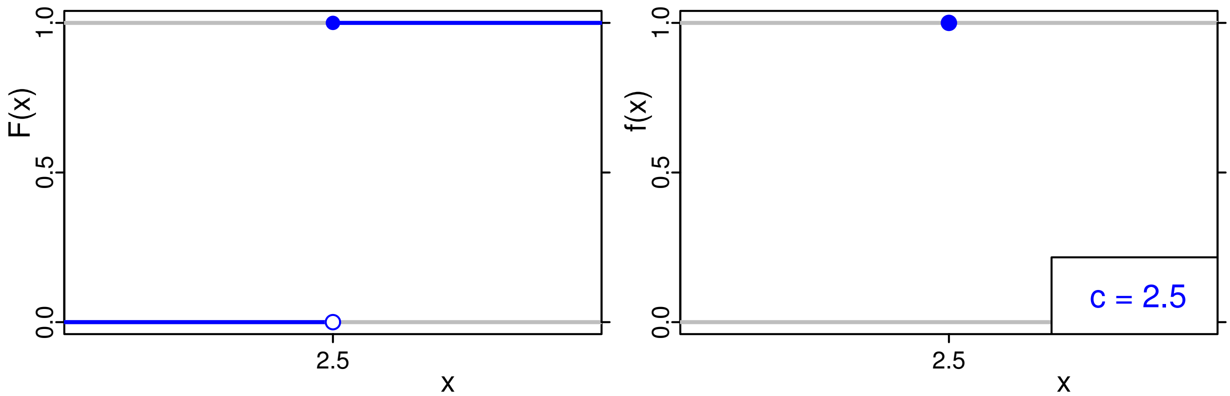


Figura 2.3. [izquierda] Gráfica de la función de distribución de una variable aleatoria constante: $X = c = 2.5$. [derecha] Respectiva gráfica de función de densidad.

Distribución Bernoulli. Considere un ensayo de Bernoulli, de modo que el espacio muestral tiene dos eventos simples, simplificados como fracaso o éxito, 0 ó 1. Véase la sección 1.6. Defina la variable aleatoria X como

$$X(\omega) = \begin{cases} 1 & \omega = \text{éxito} \\ 0 & \omega = \text{fracaso.} \end{cases}$$

La probabilidad de éxito o fracaso se establece según el contexto del fenómeno de estudio, según (1.40)-(1.41):

$$P(E) = p \quad \text{y} \quad P(F) = 1 - p; \quad \text{con} \quad 0 \leq p \leq 1.$$

En la siguiente tabla se verifica formalmente el estatus de X como variable aleatoria.

$x \in \mathbf{R}$	$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$	$F(x) = P(X \leq x)$
$x < 0$	$\varnothing$	0
$0 \leq x < 1$	$\{X = 0\} = \{F\}$	$1 - p$
$x \geq 1$	$\{X = 0\} \cup \{X = 1\} = \{F, E\} = \Omega$	1

El conjunto de valores posibles $X = 0, 1$, se define con los dos saltos de la función de distribución:

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - p & 0 \leq x < 1 \\ 1 & x \geq 1. \end{cases} \tag{2.15}$$

La respectiva función de densidad es

$$\begin{aligned} f(x) = P(X = x) &= \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases} \\ &= p^x (1 - p)^{1-x}; \quad \text{para} \quad x = 0, 1 \end{aligned} \tag{2.16}$$

Recuerde que el conjunto de valores posibles de la variable aleatoria X se determina por los puntos donde la función de densidad es positiva o, de manera equivalente, de los puntos de discontinuidad de la función de distribución $F(x)$. En este caso, $X = 0, 1$. El modelo de probabilidad (2.15), o su equivalente (2.16) es del tipo paramétrico, donde el *parámetro* p indexa a la familia de distribuciones de probabilidad Bernoulli. Cada $p \in [0, 1]$ representa a un único miembro de dicha familia, y viceversa. Se denota como

$$X \sim \mathrm{Ber}(p); \quad p \in [0, 1].$$

La Figura 2.4[arriba] muestra tres ejemplos de gráficas de la función de distribución $F(x)$, junto [abajo] con las respectivas gráficas de la función de densidad $f(x)$.

1. En el primer ejemplo, al apostar a los números rojos en un juego de la ruleta, la probabilidad de ganar es $p = 18/38 = 0.4737$. En ambas gráficas, se aprecia que la variable aleatoria X sólo toma los valores 0 ó 1. Como $p \approx 0.5$, hay mucha incertidumbre del resultado del experimento. Lo que dificulta una predicción del mismo. Sin embargo, es un poco más factible el fracaso que el éxito. Cuando se juega sólo una vez, la abertura entre p y $1-p$ tiene poco impacto negativo en el apostador. Sin embargo, esta abertura es significativamente relevante a largo plazo, cuando se juega a la ruleta muchas veces; en perjuicio del apostador y a favor de la casa. Desarrolle el Ejercicio 2.3, con $p = 18/38$.

2. El segundo ejemplo, al lanzar un dado, la probabilidad de que aparezca el número cuatro en la cara superior es $p = 1/6 = 0.1667 = 16.67\,\%$. En este caso, la población es más homogénea que la del ejemplo de la ruleta, en el sentido de que el fracaso es mucho más factible el éxito. Por lo cual, la predicción del experimento es: fracaso.

3. Para el último ejemplo, intercambie el rol de éxito y fracaso del inciso anterior, en donde éxito significa que la cara superior no es un cuatro, con probabilidad $p = 5/6 = 0.8333 = 83.33\,\%$. Aquí el éxito es altamente probable.

Distribución Binomial. Se dice que una variable aleatoria X tiene distribución *binomial*, de parámetros $n \geq 1$ y $0 \leq p \leq 1$, si su función de densidad es distribución discreta!binomial

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, \ldots, n. \tag{2.17}$$

La correspondiente función de distribución no tiene una forma cerrada:

$$F(x) = P(X \leq x) = P(X \leq \lfloor x \rfloor) = \begin{cases} 0 & x < 0 \\ \displaystyle\sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & 0 \leq x \leq n \\ 1 & x \geq n. \end{cases}$$

Aquí $\lfloor x \rfloor$ denota la *parte entera* de $x \in \mathbf{R}$; véase expresión (12) del Apéndice. Se denota como

$$X \sim \mathrm{bin}(n, p).$$

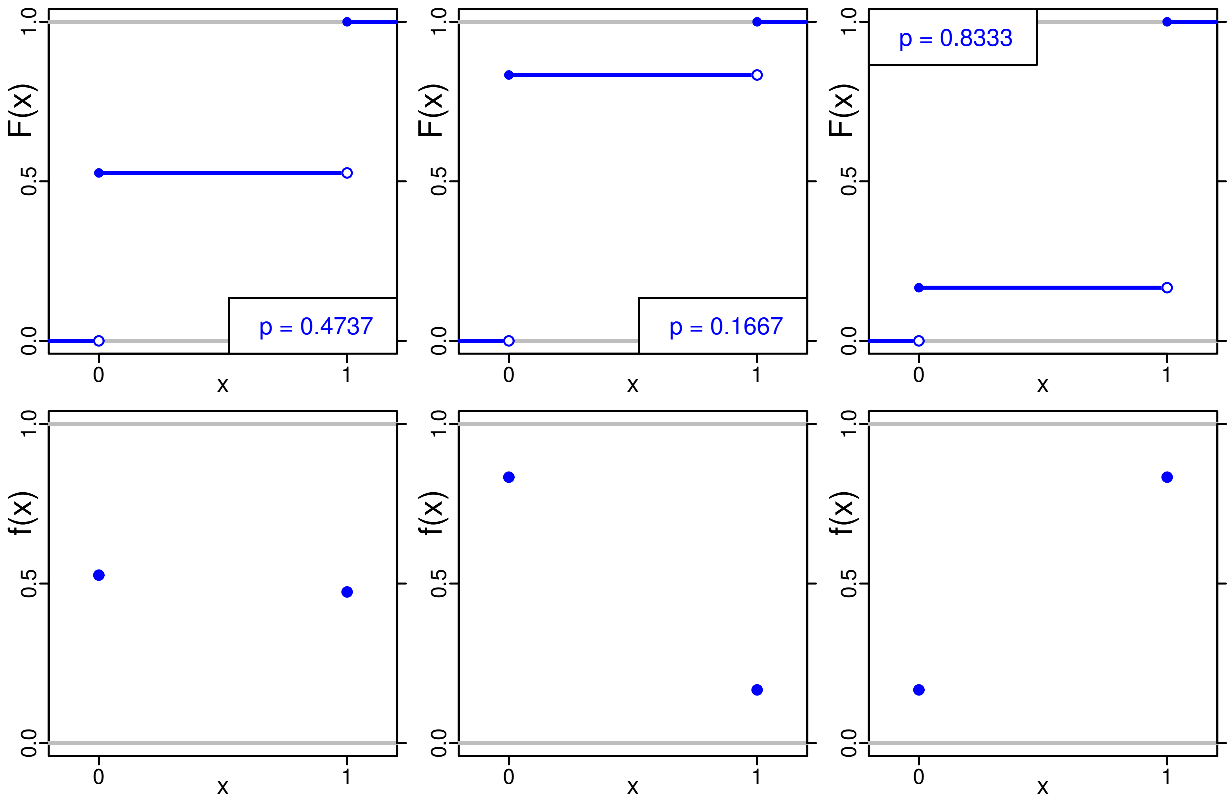


Figura 2.4. [arriba] Ejemplos de gráficas de funciones de distribución Bernoulli. [abajo] Respectivas funciones de densidad.

El caso particular $n = 1$ es de la distribución Bernoulli de parámetro p:

$$X \sim \text{Ber}(p) \sim \text{bin}(n = 1, p).$$

La génesis de la distribución binomial es la de la suma de n variables aleatorias *independientes* de distribución Bernoulli:

$$\begin{aligned} X &: \text{número de éxitos en } n \text{ ensayos de Bernoulli} \\ &= X_1 + \cdots + X_n, \end{aligned}$$

donde $X_1, \ldots, X_n$, son n variables aleatorias independientes de distribución Bernoulli de parámetro p:

$$X_i(\omega) = \begin{cases} 0 & \omega_i = \text{fracaso} \\ 1 & \omega_i = \text{éxito.} \end{cases}$$

El conjunto de valores posibles es

$$X = 0, 1, \ldots, n.$$

La independencia de las variables aleatorias se obtiene de la independencia entre los resultados de los ensayos de Bernoulli. Recuerde que, por construcción, el resultado en un ensayo específico no mejora o empeora la posibilidad de los resultados de los demás ensayos y viceversa. El tema de independencia de variables aleatorias se aborda en (2.46), más adelante.

Se han desarrollado dos ejemplos de la distribución binomial. En los ejemplos 1.6.2 y 2.1.1, se desarrolla la distribución de la variable aleatoria X, del número de rojos en una sesión de $n = 5$ el juegos de la ruleta. Así que $X \sim \text{bin}(n = 5, p = 0.4737)$. La gráfica de sus funciones de densidad (2.1) y de distribución (2.2), se muestran en la Figura 2.1. Por otro lado, el Ejemplo 1.6.3 describe de manera implícita el caso $X \sim \text{bin}(n = 5, p = 0.003)$, donde X denota el número de cilindros de concreto que fallan de un lote de $n = 7$.

Se deducirá (2.17) para el caso de $n = 10$ ensayos de Bernoulli (éxito-fracaso). En este caso, el espacio muestral es

$$\begin{aligned} \Omega &= \{(\omega_1, \ldots, \omega_{10}) : \omega_i = \text{fracaso ó éxito}\} \\ &= \{(\omega_1, \ldots, \omega_{10}) : \omega_i = F, E\} \\ &= \{(\omega_1, \ldots, \omega_{10}) : \omega_i = 0, 1\}. \end{aligned}$$

Defina la variable aleatoria

$$X(\omega) = \text{número de éxitos en 10 ensayos} = \sum_{i=1}^{10} \omega_i;$$

para cada $\omega = (\omega_1, \ldots, \omega_{10}) \in \Omega$. La Tabla 2.5 muestran algunos eventos simples del experimento, con su correspondiente valor $X(\omega)$. Recuerde que el espacio muestral Ω tiene $2^{10} = 1024$ diferentes eventos simples; véase la Sección 1.4.1. En contraste, la variable aleatoria X tiene sólo 11 valores posibles:

$$X = 0, 1, \ldots, 10.$$

Las probabilidades de los valores posibles de X se obtienen como sigue.

$$\begin{aligned}
P(X=0) &= P(FFFFFFFFFF) = (1-p)^{10} = \binom{10}{0} p^0 (1-p)^{10}, \\
P(X=1) &= P(EFFFFFFFFF) + \cdots + P(FFFFFFFFFE) \\
&= 10p(1-p)^9 = \binom{10}{1} p^1 (1-p)^9.
\end{aligned}$$

Para el evento $\{X=2\}$, haya dos letras E en la sucesión de 10 ensayos, que se consigue de $\binom{10}{2}$ diferentes maneras. Por lo que

$$P(X=2) = \binom{10}{2} P(EEFFFFFFFF) = \binom{10}{2} p^2 (1-p)^8 = 45p^2(1-p)^8.$$

Así mismo

$$\begin{aligned}
P(X=3) &= \binom{10}{3} P(EEEFFFFFFF) = \binom{10}{3} p^3 (1-p)^7 = 120p^3(1-p)^7, \\
P(X=4) &= \binom{10}{4} P(EEEEFFFFFF) = \binom{10}{4} p^4 (1-p)^6 = 210p^4(1-p)^6, \\
P(X=5) &= \binom{10}{5} P(EEEEEFFFFF) = \binom{10}{5} p^5 (1-p)^5 = 252p^5(1-p)^5, \\
P(X=6) &= \binom{10}{6} P(EEEEEEFFFF) = \binom{10}{6} p^6 (1-p)^4 = 210p^6(1-p)^4, \\
P(X=7) &= \binom{10}{7} P(EEEEEEEFFF) = \binom{10}{7} p^7 (1-p)^3 = 120p^7(1-p)^3, \\
P(X=8) &= \binom{10}{8} P(EEEEEEEEFF) = \binom{10}{8} p^8 (1-p)^2 = 45p^8(1-p)^2, \\
P(X=9) &= \binom{10}{9} P(EEEEEEEEEF) = \binom{10}{9} p^9 (1-p)^1 = 10p^9(1-p), \\
P(X=10) &= \binom{10}{10} P(EEEEEEEEEE) = \binom{10}{10} p^{10} (1-p)^0 = p^{10}.
\end{aligned}$$

Se concluye

$$\begin{aligned}
P(X=x) &= \binom{10}{x} P(\overbrace{E\cdots E}^{x}\overbrace{F\cdots F}^{10-x}) \\
&= \binom{10}{x} \overbrace{P(E)\cdots P(E)}^{x}\overbrace{P(F)\cdots P(F)}^{10-x} \\
&= \binom{10}{x} p^x (1-p)^{10-x}; \quad \text{para} \quad x = 0, 1, \ldots, 10.
\end{aligned}$$

ω	$\omega = (\omega_1, \ldots, \omega_{10})$	$X(\omega) = \sum_{i=1}^{10} \omega_i$
$FFFFFFFFFF$	0000000000	0
$EFFFFFFEFF$	1000000100	2
$EFFFFFFFFE$	1000000001	2
$EEFFFFFFFF$	1100000000	2
$EEFEFFFFFF$	1101000000	3
$FFFEEFFEFF$	0001100100	3
$EEEEEEEEEE$	1111111111	10

Tabla 2.5. Algunos eventos simples del experimento de $n = 10$ ensayos de Bernoulli, con su correspondiente suma de éxitos.

En general, si el experimento se compone de $n \geq 1$ ensayos de Bernoulli (independientes y bajo las mismas condiciones), entonces

$$\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_i = 0, 1;\ i = 1, \ldots, n\}$$

y

$$X(\omega) = \text{número de éxitos} = \sum_{i=1}^{n} \omega_i.$$

Para cada valor posible $x = 0, \ldots, n$, el evento $\{X = x\}$ representa x éxitos en n ensayos. Este evento es la unión de $\binom{n}{x}$ eventos simples, los cuales ocurren con la misma probabilidad: $p^x(1-p)^{n-x}$. Por lo tanto, se obtiene la función de densidad (2.17):

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, \ldots, n.$$

Para validar (2.17) como modelo de función de densidad, se debe satisfacer las dos propiedades del Teorema 2.1.6. La primera propiedad es inmediata al considerar los elementos positivos que involucra. Ahora bien, la segunda propiedad de dicho teorema se verifica por la fórmula del binomio (20) del Apéndice:

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1.$$

La Figura 2.5 muestra la gráfica de la función de densidad binomial de parámetros $n = 51$ y diversos valores de la probabilidad de éxito: $p = 1/6, 18/38, 0.95$. Note que la población se recorre a la derecha conforme crece p. Así mismo, dicha población es más dispersa cuando $p \approx 0.5$. distribución discreta!binomial

Distribución de Poisson. La distribución de Poisson aplica para variables aleatorias de conteo, ya sea en el tiempo o el espacio. Algunas de las variables aleatorias que se modela

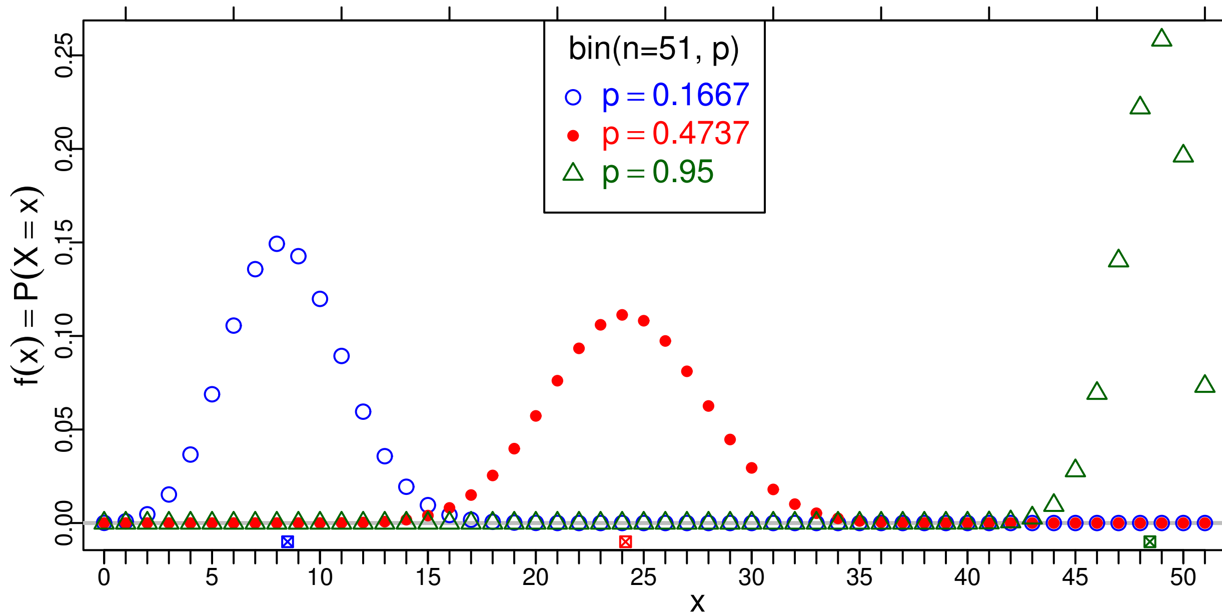


Figura 2.5. Ejemplos de gráficas de función de densidad binomial, con $n = 51$.

esta distribución de probabilidad son las siguientes

X = número de:
clientes que llegan a un servidor en una hora,
clientes que llegan a un servidor en un día,
automóviles que pasan por un crucero en un día específico,
fallas en cable de cobre en carrete de 10 km,
serpientes cascabel en la zona serrana Sierra Fría,
bacterias en una gota de agua,
errores en un libro.

Se dice que una variable aleatoria discreta X tiene distribución de *Poisson* de parámetro $\mu > 0$, si la función de densidad es

$$f(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!}; \quad \text{para} \quad x = 0, 1, 2, \ldots, \quad \text{con} \quad \mu > 0. \tag{2.18}$$

Se denota como $X \sim \text{Po}(\mu)$. El parámetro $\mu > 0$ representa la *media* de la población. La gráfica de la función de densidad se muestra en la Figura 2.6, para diversos valores de μ. Observe que tanto el centro de la población como su dispersión, crecen conforme lo hace el parámetro μ. Ahora se verificará que la función de densidad Poisson (2.18) satisface las dos propiedades características del Teorema 2.1.6. La primera propiedad es trivial, pues $f(x)$ es no negativa. Así mismo, la segunda propiedad se verifica por lo siguiente.

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} e^{-\mu}\frac{\mu^x}{x!} = e^{-\mu}\sum_{x=0}^{\infty}\frac{\mu^x}{x!} = e^{-\mu}e^{\mu} = e^{-\mu+\mu} = e^0 = 1.$$

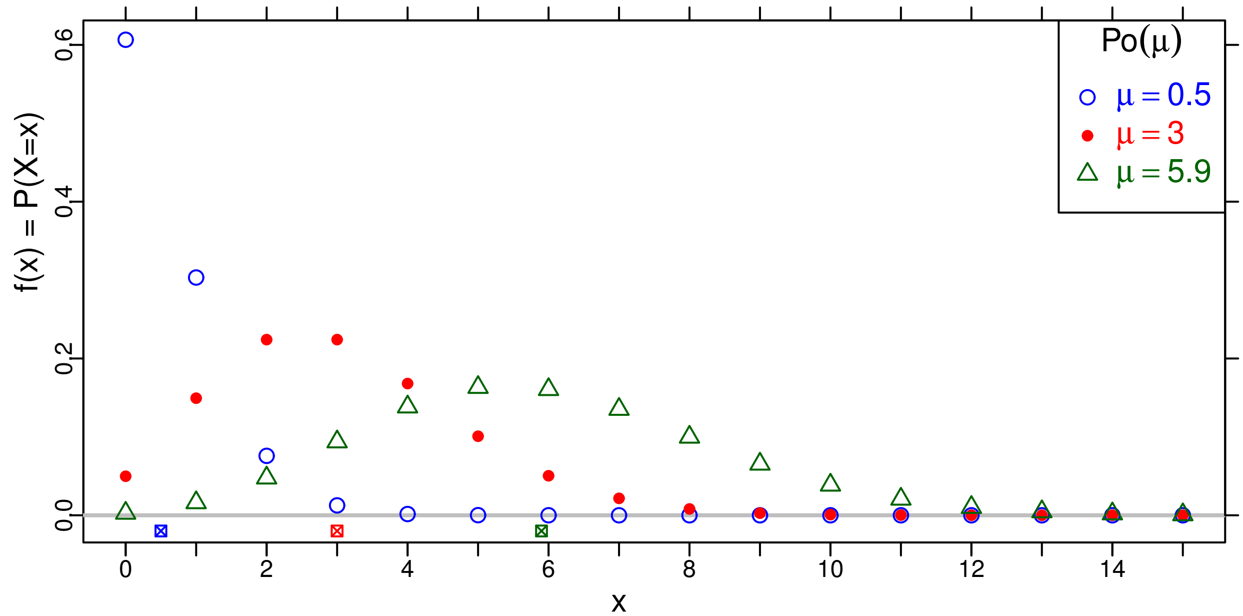


Figura 2.6. Ejemplos de gráficas de función de densidad de Poisson.

Por último, la correspondiente función de distribución no tiene una forma cerrada. La Tabla 2.4 muestra la función de distribución de Poisson con media $\mu = 2.5$. Su gráfica se muestra en la Figura 2.2[arriba].

Por ejemplo, asuma que una compañía constructora de obra civil tiene una media de $\mu = 5.4$ clientes por año. La empresa es financieramente funcional si en un año tiene cuando menos 3 proyectos. Así mismo, esta genera utilidades a partir de cuatro clientes por año. Calcule la probabilidad de los siguientes eventos del año 2030:

1. La compañía tiene déficit financiero. Lo que implica una capitalización o pedir prestado.

2. La empresa queda tablas, en donde no hay utilidades pero tampoco deudas.

3. La compañía presume utilidades, las cuales se canalizan para motivar a los trabajadores de la empresa, comprar equipo o invertir en activos financieros.

Defina la variable aleatoria X del número de clientes que tendrá la empresa en el año 2030. Su conjunto de valores posibles es $X = 0, 1, 2, \ldots$ Un modelo de probabilidad candidato para esta variable aleatoria es la distribución de Poisson. Se descartan las distribuciones uniforme discreta, binomial o geométrica. Por lo tanto, $X \sim \text{Po}(\mu = 5.4)$. Su función de densidad es

$$f(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!} = e^{-5.4}\frac{5.4^x}{x!}; \quad \text{para} \quad x = 0, 1, 2, \ldots$$

El primer evento de interés es haya déficit:

$$\{X < 3\} = \{X \leq 2\} = \{X = 0\} \cup \{X = 1\} \cup \{X = 2\}.$$

Por lo que

$$\begin{aligned} P(X \leq 2) &= F(2) = f(0) + f(1) + f(2) \\ &= e^{-5.4}\frac{5.4^0}{0!} + e^{-5.4}\frac{5.4^1}{1!} + e^{-5.4}\frac{5.4^2}{2!} \\ &= e^{-5.4}\left(1 + 5.4 + \frac{5.4^2}{2}\right) \\ &= 0.0948 = 9.48\,\%. \end{aligned}$$

El segundo evento de interés es que no haya pérdidas ni ganancias: $\{X = 3\}$. Así

$$P(X = 3) = f(3) = e^{-5.4}\frac{5.4^3}{3!} = 0.1185 = 11.85\,\%.$$

Este evento que es más probable que el anterior. Por último, es más factible que en el año 2030 haya utilidad, como lo indica su probabilidad:

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - F(3) \\ &= 1 - [f(0) + f(1) + f(2) + f(3)] \\ &= 1 - e^{-5.4}\left(1 + 5.4 + \frac{5.4^2}{2} + \frac{5.4^3}{3!}\right) = 1 - 0.2133 \\ &= 0.7867 = 78.67\,\%. \end{aligned}$$

Ahora se describirá otro ejemplo. Asuma que la duración en horas enteras de la batería de un teléfono celular tiene distribución de Poisson de media $\mu = 7.3$h. Considere que la jornada típica de un usuario de celular es de 12h. Obtenga la probabilidad de que la persona se quede sin carga de celular antes de concluir su jornada de uso. Aquí la variable aleatoria de interés se modela con la distribución de Poisson

$$\begin{aligned} X :&\ \text{duración en horas enteras de la carga del teléfono celular} \\ \sim&\ \text{Po}(\mu = 7.3). \end{aligned}$$

La función de densidad es

$$f(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!} = e^{-7.3}\frac{7.3^x}{x!}; \quad \text{para} \quad x = 0, 1, 2, \ldots$$

El evento de interés es $\{X < 12\} = \{X \leq 11\}$. Así

$$\begin{aligned} P(X \leq 11) &= F(11) = f(0) + f(1) + \cdots + f(11) \\ &= e^{-7.3}\left(1 + 7.3 + \frac{7.3^2}{2} + \cdots + \frac{7.3^{11}}{11!}\right) \\ &= 0.9319. \end{aligned}$$

Lo que significa que el usuario se quedará sin carga la mayoría de los días. Esta información es útil en la toma de decisiones del usuario de celular: (a) cambiar de aparato a uno mejor, (b) reducir su uso durante el día, (c) portar un cargador o batería externa, (d) usarlo hasta que se agote la carga, entre otras opciones. Por otro lado, los diseñadores de celulares procurarán nuevos modelos de baterías que preserven una mayor carga y se adapten a las necesidades de sus clientes.

Distribución geométrica. Considere el experimento de una secuencia de ensayos de Bernoulli, independientes con resultados éxito o fracaso en cada ensayo. Registre la variable aleatoria

$$X : \text{número de ensayos hasta el primer éxito.}$$

Aquí se asume que el experimento se detiene o deja de observar hasta que ocurra el primer éxito. El conjunto de valores posibles de esta variable aleatoria es

$$X = 1, 2, 3, \ldots$$

La función de densidad *geométrica* es

$$f(x) = P(X = x) = p(1-p)^{x-1}; \quad x = 1, 2, 3, \ldots, \tag{2.19}$$

donde $p = P(E) \in (0, 1)$. Se denota como

$$X \sim \text{geo}(p); \quad \text{con} \quad 0 < p < 1.$$

La Tabla 2.6 muestra los eventos simples del espacio muestral, junto con los respectivos valores de la variable aleatoria X y sus probabilidades. La distribución geométrica es un modelo de probabilidad válido si su función de densidad (2.19) es positiva y suma uno. De hecho

$$\begin{aligned}
\sum_{x=0}^{\infty} f(x) = \sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} p(1-p)^{x-1} \\
&= p\sum_{x=1}^{\infty}(1-p)^{x-1} = p\sum_{z=0}^{\infty}(1-p)^{z} \\
&= p\frac{1}{1-(1-p)} = \frac{p}{p} \\
&= 1.
\end{aligned}$$

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} p(1-p)^{x-1} = p\sum_{z=0}^{\infty}(1-p)^{z} = \frac{p}{1-(1-p)} = 1.$$

En la penúltima igualdad se aplicó la serie geométrica (22) del Apéndice. Por lo que se refiere a la función de distribución $F(x)$, esta se evaluará en el conjunto de valores posibles, que son los puntos donde $F(x)$ tiene una discontinuidad o salto.

$$\begin{aligned}
F(x) = P(X \le x) &= \sum_{y=0}^{x} f(y) \\
&= \sum_{y=1}^{x} f(y) = \sum_{y=1}^{x} p(1-p)^{y-1} \\
&= p\sum_{z=0}^{x-1}(1-p)^{z} = p\frac{1-(1-p)^{(x-1)+1}}{1-(1-p)} \\
&= 1-(1-p)^{x}; \quad \text{para} \quad x = 1, 2, \ldots
\end{aligned}$$

evento simple ω	número de ensayos $X(\omega)$	$P(X=x)$	número de fracasos $Y(\omega)=X(\omega)-1$	$P(Y=y)$
E	1	p	0	p
FE	2	$(1-p)p$	1	$(1-p)p$
FFE	3	$(1-p)^2p$	2	$(1-p)^2p$
$FFFE$	4	$(1-p)^3p$	3	$(1-p)^3p$
$FFFFE$	5	$(1-p)^4p$	4	$(1-p)^4p$
$\vdots$	0			
$\underbrace{F\cdots F}_{x-1}E$	x	$(1-p)^{x-1}p$	$y=x-1$	$(1-p)^yp$
$\vdots$				

Tabla 2.6. Conjunto de valores posibles de una variable aleatoria de distribución geométrica; para las variables aleatorias número de ensayos X y número de fracasos Y.

Por tanto, la función de distribución $F(x)$ evaluada en la recta real es

$$F(x)=P(X\leq x)=\begin{cases}0 & x<1\\ 1-(1-p)^{\lfloor x\rfloor} & x\geq 1.\end{cases} \tag{2.20}$$

Esta función es creciente, continua por la derecha y satisface

$$F(-\infty)=0 \quad \text{y} \quad F(\infty)=1.$$

La primera igualdad es inmediata. Para la segunda, se tiene

$$F(\infty)=\lim_{x\to\infty}F(x)=\lim_{x\to\infty}[1-(1-p)^x]=1-\lim_{x\to\infty}(1-p)^x=1-0=1.$$

Por otro lado, se aplicará (2.11) para deducir la función de densidad (2.19) en términos de su función de distribución (2.20). Para $x=1,2,\ldots$, se tiene

$$\begin{aligned}
f(x)=P(X=x)&=F(x)-F(x-)\\
&=P(X\leq x)-P(X<x)=P(X\leq x)-P(X\leq x-1)\\
&=F(x)-F(x-1)=[1-(1-p)^x]-[1-(1-p)^{x-1}]\\
&=(1-p)^{x-1}-(1-p)^x=(1-p)^{x-1}[1-(1-p)]\\
&=p(1-p)^{x-1}.
\end{aligned}$$

La Figura 2.7 muestra tres ejemplos de la gráfica de la función de densidad geométrica: $p=0.2$; punto hueco azul, $p=0.5$; punto rojo y $p=0.9$; triángulo verde. Note que, a mayor valor de p, la distribución tiende a concentrarse en valores chicos de x.

Una versión alternativa de la distribución geométrica es el de la variable aleatoria

$$Y: \text{número de fracasos hasta el primer éxito.}$$

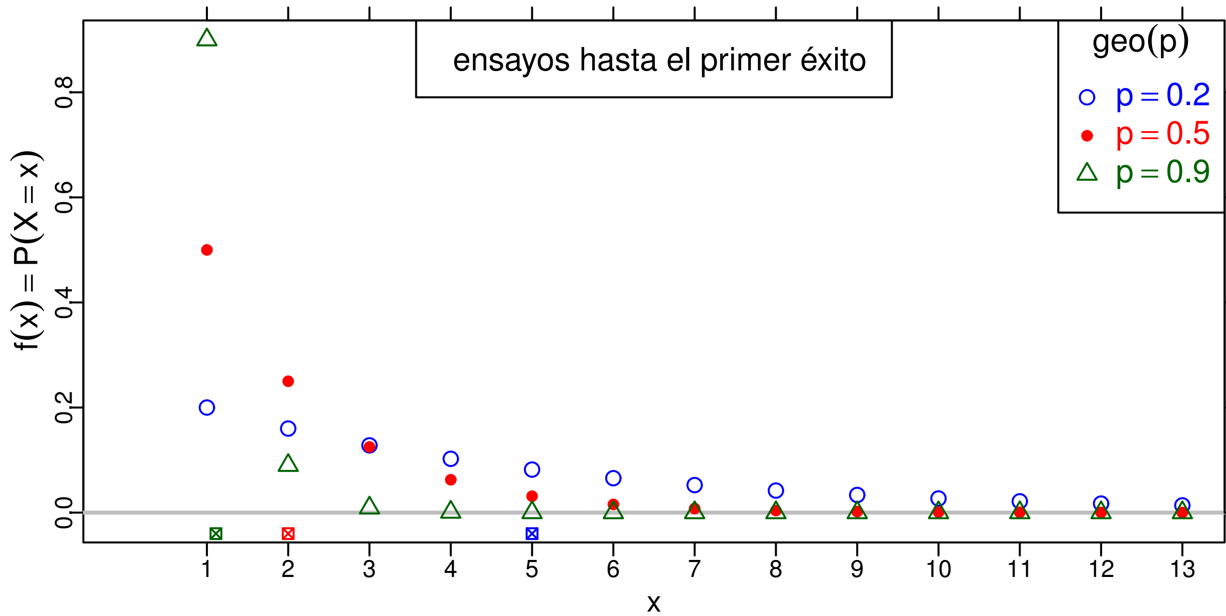


Figura 2.7. Ejemplos de gráficas de función de densidad geométrica.

Entonces

$$Y = X - 1 \quad \text{y} \quad X = Y + 1.$$

El conjuntos de valores posibles de la variable aleatoria Y, junto con sus probabilidades respectivas, se muestra en la Tabla 2.6. Note que tales probabilidades coinciden con las obtenidas para los valores posibles de la variable aleatoria X. De hecho, la función de densidad de Y se deduce como sigue

$$\begin{aligned} P(Y = y) &= P(X - 1 = y) = P(X = y + 1) \\ &= p(1 - p)^{(y+1)-1} = p(1 - p)^y; \quad \text{para} \quad y = 0, 1, 2, \ldots \end{aligned} \tag{2.21}$$

A la distribución de la variable aleatoria Y también se le dice *geométrica* de parámetro p. Por el Ejercicio 2.11.5, se obtienen la funciones de distribución y de *supervivencia*. Cabe aclarar que, si bien comparten el mismo nombre, las distribuciones de X y Y son diferentes. Para evitar alguna confusión, hay que distinguir a cual versión corresponde. La Figura 2.8[arriba] muestra la gráfica de la función de densidad geométrica de parámetro $p = 0.3$, del número de fracasos hasta el primer éxito, junto con su respectiva función de distribución [abajo].

Distribución binomial negativa. La distribución binomial negativa es una extensión de la distribución geométrica, en donde se requiere de $r \geq 1$ éxitos para detener la secuencia de ensayos de Bernoulli. En particular, la distribución geométrica corresponde a $r = 1$ éxitos.

Considere una secuencia de ensayos de Bernoulli (éxito-fracaso), con probabilidad de éxito $p = P(E) > 0$. Sea $r \geq 1$ el número de éxitos requerido y defina la variable aleatoria

$$X : \text{número de ensayos hasta el } r\text{-ésimo éxito}.$$

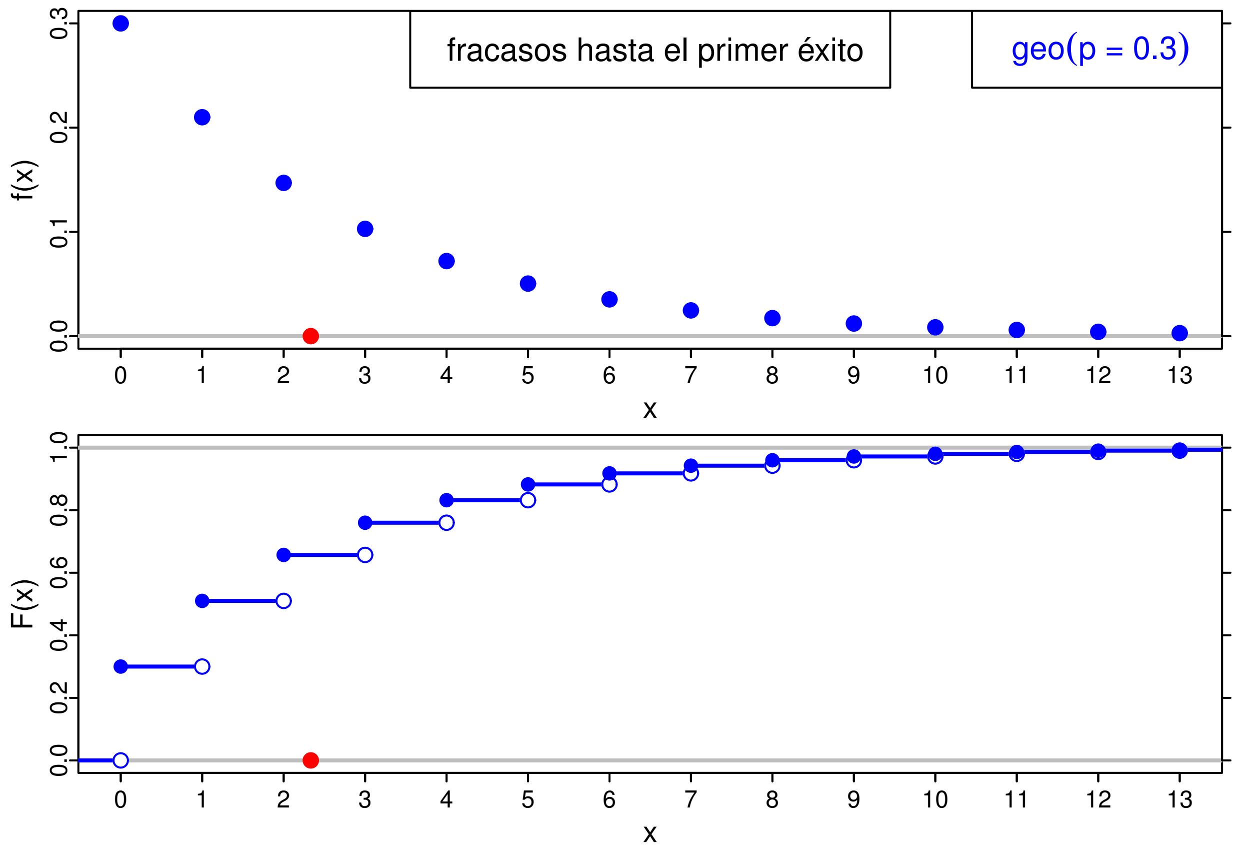


Figura 2.8. [arriba] Gráfica de función de densidad geométrica, del número de fracasos hasta el primer éxito. [abajo] Respectiva función de distribución.

El conjunto de valores posibles es

$$X = r, r+1, r+2, \ldots$$

La función de densidad es

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}; \quad \text{para} \quad x = r, r+1, r+2, \ldots$$

La suma de r variables aleatorias independientes de distribución geométrica de parámetro p tiene distribución binomial negativa:

$$X = X_1 + \cdots + X_r \sim \text{binneg}(r, p). \tag{2.22}$$

Si por ejemplo se requieren $r = 3$ éxitos, entonces X_1 es el número de ensayos hasta lograr el primer éxito. A partir de ahí, se requieren X_2 ensayos hasta que llegué el segundo ensayo. Al contabilizar de nuevo, hay que esperar otros X_3 ensayos para el tercer éxito. Las variables aleatorias X_1, X_2 y X_3 tienen la misma distribución geométrica de parámetro p, además de son independientes. En este sentido, la magnitud de X_1 no influye en los valores de X_2 y X_3 y, viceversa. Entonces

$$X = X_1 + X_2 + X_3 \sim \text{binneg}(r = 3, p). \tag{2.23}$$

Por ejemplo si $r = 5$, entonces

$$\begin{aligned} X &: \text{ número de ensayos hasta el quinto éxito} \\ &= 5, 6, 7, \ldots \end{aligned}$$

Se requiere al menos 5 ensayos para alcanzar el mismo número de éxitos. Considere el evento

$$\{X = 5\} = \{EEEEE\}.$$

Entonces

$$P(X = 5) = P(EEEEE) = P(E)P(E)P(E)P(E)P(E) = p^5.$$

Así mismo, el eventos $\{X = 6\}$ es la unión de 5 eventos simples equiprobables:

$$\{X = 6\} = \{FEEEEE, EFEEEE, EEFEEE, EEEFEE, EEEEFE\}.$$

Note que en estos eventos simples el único fracaso se encuentra en cualquiera de las primeras cinco posiciones, mientras que la sexta y última posición debe ser éxito. Así, el total de eventos simples del evento $\{X = 6\}$ es

$$5 = \binom{5}{4} = \binom{6-1}{5-1}.$$

Por lo que

$$\begin{aligned} P(X = 6) &= 5P(FEEEEE) = 5P(F)P(E)P(E)P(E)P(E)P(E) \\ &= 5(1-p)p^5 = \binom{6-1}{5-1} p^5 (1-p). \end{aligned}$$

De manera similar, se obtiene

$$\begin{aligned}
&\{X = 7\}\\
&= \{FFEEEEE, FEFEEEE, FEEFEEE, FEEEFEE, FEEEEFE,\\
&\quad EFFEEEE, EFEFEEE, EFEEFEE, EFEEEFE, EEFFEEE,\\
&\quad EEFEFEE, EEFEEFE, EEEFFEE, EEEFEFE, EEEEFFE\}.
\end{aligned}$$

El evento $\{X = 7\}$ requiere que el séptimo y último ensayo sea un éxito E. Además, en los primeros 6 ensayos hay 2 fracasos, ubicados de 15 maneras diferentes:

$$15 = \binom{6}{4} = \binom{7-1}{5-1}.$$

Por lo que

$$\begin{aligned}
P(X = 7) &= 15P(FFEEEEE) = 15[P(F)]^2[P(E)]^5\\
&= 15(1-p)^2p^5 = \binom{7-1}{5-1}p^5(1-p)^2.
\end{aligned}$$

Así, para $x = 5, 6, 7, \ldots,$ se tiene

$$\begin{aligned}
P(X = x) &= \binom{x-1}{4}P(\overbrace{F\cdots F}^{x-5}EEEEE)\\
&= \binom{x-1}{5-1}[P(F)]^{x-5}[P(E)]^5\\
&= \binom{x-1}{5-1}p^5(1-p)^{x-5}.
\end{aligned}$$

2.3. Esperanza de una variable aleatoria discreta

En esta sección se define el concepto de esperanza de una variable aleatoria discreta. Se describen sus principales propiedades, así como algunos ejemplos.

Sea X una variable aleatoria discreta con función de densidad $f(x) = P(X = x)$; para x en el conjunto de valores posibles $\mathcal{X}$; ya sea finito o infinito numerable. Si

$$\sum_x |x|f(x) < \infty, \tag{2.24}$$

donde la suma es sobre los elementos $x \in \mathcal{X}$, entonces la *esperanza*, *valor esperado* o *media poblacional* de la variable aleatoria X es

$$\mu = \mathrm{E}\,X = \sum_x xf(x). \tag{2.25}$$

Si en particular, el conjunto de valores posibles son los enteros no negativos, entonces

$$\mu = \mathrm{E}\,X = \sum_{x=0}^{\infty} xf(x).$$

El valor esperado $\mu \in \mathbf{R}$ es el centro de la población. Sus individuos se encuentran alrededor de dicho punto. Este número real resume, integra y predice a su variable aleatoria. En el contexto de Física, se le conoce como *centro de masa*. En cambio, si diverge la suma (2.24), se dice que X no tiene esperanza. En el caso que X sea una variable no negativa sin esperanza, se escribe $EX = \infty$.

Teorema 2.3.1. Sea X una variable aleatoria discreta con función de densidad $f(x) = P(X = x)$. Considere la función real $g(x)$, tal que

$$\sum_x |g(x)| f(x) < \infty.$$

Entonces

$$\mathrm{E}\, g(X) = \sum_x g(x) f(x). \tag{2.26}$$

Demostración. Sea el conjunto de valores posibles de la variable aleatoria X

$$\mathcal{X} = \{x \in \mathbf{R} : f(x) > 0\}.$$

La función real $Y = g(X)$ también es una variable aleatoria discreta, con conjunto de valores posibles

$$\mathcal{Y} = g(\mathcal{X}) = \{y \in \mathbf{R} : f_Y(y) > 0\} = \{y = g(x) : x \in \mathcal{X}\}.$$

Para $y \in \mathcal{Y}$, se tiene

$$\{Y = y\} = \{g(X) = y\} = \{X \in A_y\} = \bigcup_{x \in A_y} \{X = x\},$$

donde $A_y = \{x \in \mathcal{X} : g(x) = y\}$. Así

$$f_Y(y) = P(Y = y) = P(X \in A_y) = \sum_{x \in A_y} P(X = x) = \sum_{x \in A_y} f(x).$$

Además, la familia de conjuntos $\{A_y\}_{y \in \mathcal{Y}}$, es una partición de $\mathcal{X}$, pues son excluyentes y

$$\mathcal{X} = \bigcup_{y \in \mathcal{Y}} A_y.$$

Entonces

$$\begin{aligned}
\mathrm{E}\, Y &= \sum_y y f_Y(y) = \sum_y y \sum_{x \in A_y} f(x) \\
&= \sum_y \sum_{x \in A_y} y f(x) = \sum_y \sum_{x \in A_y} g(x) f(x) \\
&= \sum_x g(x) f(x) < \infty.
\end{aligned}$$

$\square$

Por (2.26), hay dos fórmulas de la esperanza de la variable aleatoria $Y = g(X)$:

$$\operatorname{E} Y = \sum_y y f_Y(y) = \sum_x g(x) f(x).$$

Por lo regular, la última es más práctica, pues omite el cálculo de la función de densidad de Y. Por ejemplo

$$\begin{aligned}
\operatorname{E} X^2 &= \sum_x x^2 f(x), \\
\operatorname{E}(X-\mu)^2 &= \sum_x (x-\mu)^2 f(x), \\
\operatorname{E} e^X &= \sum_x e^x f(x), \\
\operatorname{E} \log X &= \sum_x \log x f(x).
\end{aligned}$$

El último ejemplo aplica sólo si $\log X$ está bien definida, es decir, si la variable aleatoria X es positiva: $P(X > 0) = 1$. Los primeros dos ejemplos están relacionados con la varianza, que a continuación se define. Si

$$\sum_x x^2 f(x) < \infty, \tag{2.27}$$

la *varianza* o *varianza poblacional* de la variable aleatoria discreta X es

$$\sigma^2 = \operatorname{V} X = \operatorname{E}(X-\mu)^2 = \sum_x (x-\mu)^2 f(x), \tag{2.28}$$

donde $\mu = \operatorname{E} X \in \mathbf{R}$ es la media poblacional. La varianza es una suma ponderada de los errores al cuadrado; errores respecto de la media μ. Este es un indicador de la *dispersión* de la población. A mayor varianza mayor será la diversidad entre los individuos de la misma. En cambio, los individuos serán más parecidos entre sí cuando $\sigma^2 \approx 0$. En cualquier caso, la varianza es no negativa: $\sigma^2 \geq 0$. El caso $\sigma^2 = 0$ significa que la variable aleatoria X es constante, es decir, cualquier resultado del experimento $\omega \in \Omega$, implica $X(\omega) = \mu$. Note la equivalencia de la convergencia de las series (2.27) y (2.28):

$$\begin{aligned}
\sum_x (x-\mu)^2 f(x) &= \sum_x (x^2 - 2\mu x + \mu^2) f(x) \\
&= \sum_x x^2 f(x) + \sum_x (-2\mu) x f(x) + \sum_x \mu^2 f(x) \\
&= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x) \\
&= \sum_x x^2 f(x) - 2\mu \cdot \mu + \mu^2 \cdot 1 \\
&= \sum_x x^2 f(x) - \mu^2.
\end{aligned}$$

Por lo tanto

$$\sigma^2 = \mathrm{V}\, X = \mathrm{E}(X-\mu)^2 = \mathrm{E}\, X^2 - \mu^2 = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2. \tag{2.29}$$

Otra medida de dispersión es la *desviación estándar*:

$$\sigma = \mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\, X} = \sqrt{\sigma^2}. \tag{2.30}$$

Esta medida de dispersión comparte la unidad de medida de la variable aleatoria X, así como de su media μ. A continuación se mostrarán algunos ejemplos de distribuciones de probabilidad.

Distribución uniforme discreta. Una población tiene distribución *uniforme discreta* en $\{1,\ldots,N\}$, si su función de densidad es

$$f(x) = P(X=x) = \frac{1}{N}; \quad x = 1,\ldots,N, \tag{2.31}$$

con $N \geq 1$. Se denota como $X \sim \mathrm{U}(1,\ldots,N)$. La media y varianza de la variable aleatoria X son, respectivamente

$$\mu = \mathrm{E}\, X = \frac{N+1}{2} \quad \text{y} \quad \sigma^2 = \mathrm{V}\, X = \frac{N^2-1}{12}. \tag{2.32}$$

Tanto la media como la dispersión de esta distribución son cada vez mayores conforme N sea más grande. Su deducción se describe a continuación.

$$\begin{aligned} \mathrm{E}\, X &= \sum_{x=0}^{\infty} x f(x) = \sum_{x=1}^{N} x \frac{1}{N} \\ &= \frac{1}{N}\sum_{x=1}^{N} x = \frac{1}{N}\frac{N(N+1)}{2} \\ &= \frac{N+1}{2}. \end{aligned}$$

$$\begin{aligned} \mathrm{E}\, X^2 &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=1}^{N} x^2 \frac{1}{N} \\ &= \frac{1}{N}\sum_{x=1}^{N} x^2 = \frac{1}{N}\frac{N(N+1)(2N+1)}{6} \\ &= \frac{(N+1)(2N+1)}{6}. \end{aligned}$$

Al considerar (2.29), se tiene

$$\begin{aligned} \mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{N+1}{2}\left(\frac{2N+1}{3} - \frac{N+1}{2}\right) = \frac{N+1}{2}\,\frac{2(2N+1)-3(N+1)}{6} \\ &= \frac{N+1}{2}\,\frac{N-1}{6} = \frac{N^2-1}{12}. \end{aligned}$$

La distribución uniforme discreta aplica a un sorteo de N boletos, donde X es el número del boleto ganador seleccionado al azar. Por ejemplo, en el lanzamiento de un dado, se tiene $N = 6$ y $X \sim \mathrm{U}(1, 2, 3, 4, 5, 6)$. Su función de densidad es

$$f(x) = P(X = x) = \frac{1}{N} = \frac{1}{6} = 0.1667; \quad \text{para} \quad x = 1, 2, 3, 4, 5, 6.$$

Así mismo

$$\begin{aligned} \mathrm{E}\, X &= \frac{N+1}{2} = \frac{6+1}{2} = 3.5, \\ \mathrm{V}\, X &= \frac{N^2-1}{12} = \frac{6^2-1}{12} = 2.9167, \\ \mathrm{d.\,e.}(X) &= \sqrt{\mathrm{V}\, X} = \sqrt{2.9167} = 1.7078. \end{aligned}$$

Así mismo, en una rifa de $N = 10$ personas, el número X del boleto ganador tiene distribución $X \sim \mathrm{U}(1, \ldots, 10)$. La función de densidad es es

$$f(x) = P(X = x) = \frac{1}{N} = \frac{1}{10} = 0.1; \quad \text{para} \quad x = 1, \ldots, 10.$$

En este caso, son más grandes la media, varianza y desviación estándar:

$$\begin{aligned} \mathrm{E}\, X &= \frac{N+1}{2} = \frac{10+1}{2} = 5.5, \\ \mathrm{V}\, X &= \frac{N^2-1}{12} = \frac{10^2-1}{12} = 8.25, \\ \mathrm{d.\,e.}(X) &= \sqrt{\mathrm{V}\, X} = \sqrt{8.25} = 2.8723. \end{aligned}$$

La Figura 2.9 muestra la gráfica de la función de densidad y de la función de distribución, de los ejemplos aquí descritos, de la distribución uniforme discreta $\mathrm{U}(1, \ldots, N)$, con $N = 6$ y $N = 10$.

El sorteo de la Lotería Nacional aporta un ejemplo de transformación de localización. Sea Y el número del boleto ganador del premio mayor, en una emisión de $N = 50\,000$ boletos. Cada boleto está numerado de 0 a $N - 1 = 49\,999$. Todos los boletos tienen la misma probabilidad de ganar. Entonces

$$Y \sim \mathrm{U}(0, \ldots, N-1),$$

Por el Ejercicio 2.8, con $a = 0$ y $b = N - 1 = 49999$, se verifica:

$$\begin{aligned} \mathrm{E}\, Y &= \frac{a+b}{2} = \frac{0+49999}{2} = 24\,999.5, \\ \mathrm{V}\, Y &= \frac{(b-a+1)^2-1}{12} = \frac{(49999-0+1)^2-1}{12} \\ &= \frac{50000^2-1}{12} = 208\,333\,333, \\ \mathrm{d.\,e.}(Y) &= \sqrt{\mathrm{V}\, Y} = \sqrt{208333333} = 14\,433.76. \end{aligned}$$

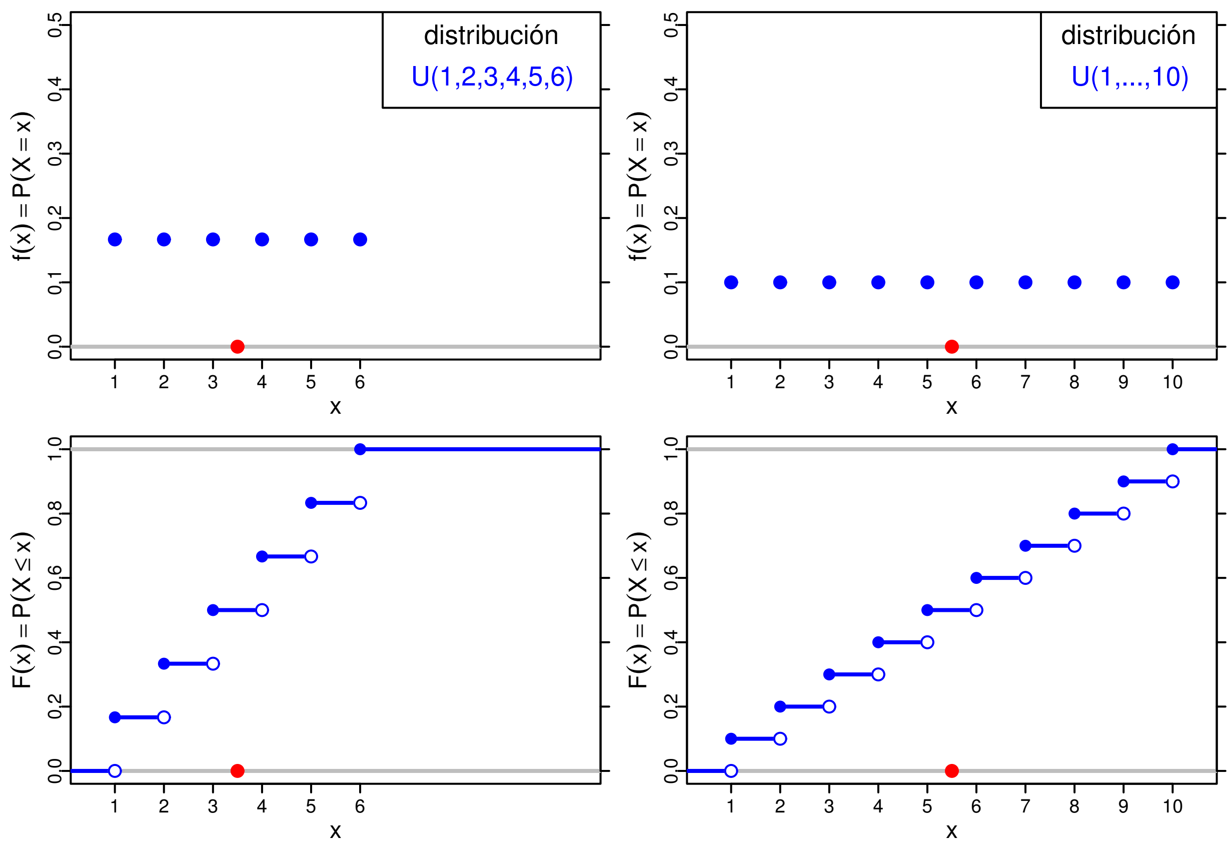


Figura 2.9. [arriba] Ejemplos de gráficas de funciones de densidad uniforme discreta. [abajo] Respectivas gráficas de funciones de distribución.

Note que $X = Y + 1$ también tiene distribución uniforme:

$$X \sim \mathrm{U}(1, \ldots, N).$$

De (2.32), se tiene

$$\begin{aligned}
\mathrm{E}\, X &= \frac{N+1}{2} = \frac{50000+1}{2} = 25000.5, \\
\mathrm{V}\, X &= \frac{N^2+1}{12} = \frac{50000^2-1}{12} = 208\,333\,333, \\
\mathrm{d.\,e.}(X) &= \sqrt{\mathrm{V}\, X} = \sqrt{208\,333\,333} = 14\,433.76.
\end{aligned}$$

Por lo tanto, se confirman los momentos de la variable aleatoria Y:

$$\begin{aligned}
\mathrm{E}\, Y &= \mathrm{E}[X-1] = 25000.5 - 1 = 24999.5 \\
\mathrm{V}\, Y &= \mathrm{V}[X-1] = \mathrm{V}\, X = 208\,333\,333.
\end{aligned}$$

A continuación se deducirán las fórmulas de la media y varianza poblacional para las distribuciones paramétricas binomial, Poisson y geométrica.

Si X es una variable aleatoria de distribución binomial, $X \sim \mathrm{bin}(n,p)$; con $n \geq 1$ y $0 \leq p \leq 1$, entonces

$$\begin{aligned}
\mathrm{E}\, X &= np, && (2.33) \\
\mathrm{V}\, X &= np(1-p), && (2.34) \\
\mathrm{d.\,e.}(X) &= \sqrt{np(1-p)}.
\end{aligned}$$

El ejemplo particular de la distribución Bernoulli ($n = 1$), implica

$$\mathrm{E}\, X = p, \quad \mathrm{V}\, X = p(1-p) \quad \text{y} \quad \mathrm{d.\,e.}(X) = \sqrt{p(1-p)}.$$

Observe que el valor esperado no requiere ser un valor posible de X. Para comprobar (2.33), recuerde la función de densidad binomial (2.17):

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad \text{para} \quad x = 0, \ldots, n.$$

El caso $n = 1$ se refiere a la distribución Bernoulli de parámetro p. La variable aleatoria X sólo toma dos valores $X = 0, 1$. Así

$$\mathrm{E}\, X = \sum_{x=0}^{\infty} x f(x) = 0P(X=0) + 1P(X=1) = 0(1-p) + 1p = p.$$

Así mismo, para la varianza, resulta

$$\mathrm{E}\, X^2 = \sum_{x=0}^{\infty} x^2 f(x) = 0^2 f(0) + 1^2 f(1) = 0(1-p) + 1p = p$$

y
$$\mathrm{V}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = p - p^2 = p(1-p).$$

Ahora bien, para $n \geq 1$, se tiene

$$\begin{aligned}
\mathrm{E}\,X &= \sum_{x=0}^{\infty} x f(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{xn(n-1)!}{x(x-1)!(n-x)!} p p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)} \\
&= np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!(n-1-z)!} p^{z} (1-p)^{n-1-z} \\
&= np \sum_{z=0}^{n-1} \binom{n-1}{z} p^{z} (1-p)^{n-1-z} \\
&= np.
\end{aligned}$$

Aquí se usó el cambio de variable $z = x - 1$. La última igualdad se debe a que $g(z) = \binom{n-1}{z} p^z (1-p)^{n-1-z}$; $z = 0, \ldots, n-1$, es la función de densidad binomial de parámetros $n-1$ y p; con lo que su suma es uno. La deducción de la varianza (2.34) es similar:

$$\begin{aligned}
\mathrm{E}\,X^2 &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x(x-1+1) \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} [x(x-1)+x] \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} + \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)!(n-x)!} p^2 p^{x-2} (1-p)^{n-x} + \mathrm{E}\,X \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!((n-2)-(x-2))!} p^{x-2} (1-p)^{(n-2)-(x-2)} + np.
\end{aligned}$$

Al considerar el cambio de variable $z = n - 2$, la última suma se simplifica como

$$\begin{aligned}\sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!((n-2)-(x-2))!} p^{x-2}(1-p)^{(n-2)-(x-2)} &= \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!} p^{z}(1-p)^{n-2-z} \\ &= \sum_{z=0}^{n-2} \binom{n-2}{z} p^{z}(1-p)^{n-2-z} \\ &= 1.\end{aligned}$$

Esta suma vale uno, pues sus términos corresponden a la función de densidad binomial de parámetros $n - 2$ y p. Así, por (2.29), se tiene

$$\begin{aligned}\mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 &= [n(n-1)p^2 + np] - (np)^2 \\ &= (n^2 - n - n^2)p^2 + np = np - np^2 \\ &= np(1-p).\end{aligned}$$

Para la distribución de Poisson, $X \sim \mathrm{Po}(\lambda)$, la media y varianza es su propio parámetro:

$$\mathrm{E}\, X = \mathrm{V}\, X = \mu > 0.$$

Así que la desviación estándar es $\mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\, X} = \sqrt{\mu}$. Por (2.18), su función de densidad es

$$f(x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!}; \quad x = 0, 1, \ldots$$

Así

$$\begin{aligned}\mathrm{E}\, X = \sum_{x=0}^{\infty} x f(x) &= \sum_{x=0}^{\infty} x e^{-\mu}\frac{\mu^x}{x!} \\ &= e^{-\mu}\sum_{x=0}^{\infty} x\frac{\mu^x}{x!} = e^{-\mu}\sum_{x=1}^{\infty}\frac{\mu\mu^{x-1}}{(x-1)!} \\ &= \mu e^{-\mu}\sum_{x=1}^{\infty}\frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu}\sum_{z=0}^{\infty}\frac{\mu^z}{z!} \\ &= \mu e^{-\mu}\cdot e^{\mu} = \mu.\end{aligned}$$

Así mismo

$$\begin{aligned}
\mathrm{E}\, X^2 &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=0}^{\infty} x^2 e^{-\mu} \frac{\mu^x}{x!} \\
&= e^{-\mu} \sum_{x=1}^{\infty} x \frac{\mu^x}{(x-1)!} = e^{-\mu} \sum_{x=1}^{\infty} (x-1+1) \frac{\mu^x}{(x-1)!} \\
&= e^{-\mu} \sum_{x=1}^{\infty} (x-1) \frac{\mu^x}{(x-1)!} + e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^x}{(x-1)!} \\
&= e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^{x-2+2}}{(x-2)!} + e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu^{x-1+1}}{(x-1)!} \\
&= e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^2 \mu^{x-2}}{(x-2)!} + e^{-\mu} \sum_{x=1}^{\infty} \frac{\mu \mu^{x-1}}{(x-1)!} \\
&= \mu^2 e^{-\mu} \sum_{z=0}^{\infty} \frac{\mu^z}{z!} + \mu e^{-\mu} \sum_{z=0}^{\infty} \frac{\mu^z}{z!} \\
&= \mu^2 e^{-\mu} e^{\mu} + \mu e^{-\mu} e^{\mu} = \mu^2 + \mu.
\end{aligned}$$

Por (2.29), resulta

$$\mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = (\mu^2 + \mu) - \mu^2 = \mu.$$

Por otro lado, si X es una variable aleatoria de distribución geométrica de parámetro $p = P(E) > 0$, $X \sim \mathrm{geo}(p)$, su media y varianza resultan

$$\mathrm{E}\, X = \frac{1}{p} \quad \text{y} \quad \mathrm{V}\, X = \frac{1-p}{p^2}. \tag{2.35}$$

La desviación estándar es

$$\mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\, X} = \frac{\sqrt{1-p}}{p}.$$

Por (2.19), su función de densidad es

$$f(x) = P(X = x) = p(1-p)^{x-1}; \quad x = 1, 2, 3, \ldots$$

Entonces

$$\begin{aligned}
\mathrm{E}\, X &= \sum_{x=0}^{\infty} x f(x) = \sum_{x=1}^{\infty} x f(x) \\
&= \sum_{x=1}^{\infty} x p (1-p)^{x-1} = p \sum_{x=0}^{\infty} x (1-p)^{x-1}.
\end{aligned}$$

El término $x(1-p)^{x-1}$ tiene forma de la derivada de una función polinomial respecto de p, incluso desde el índice $x = 0$:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p} \{(1-p)^x\} &= -x(1-p)^{x-1}, \\
-\frac{\mathrm{d}}{\mathrm{d}p} \{(1-p)^x\} &= x(1-p)^{x-1}; \quad x = 0, 1, \ldots
\end{aligned}$$

Entonces

$$\begin{aligned}\sum_{x=0}^{\infty} x(1-p)^{x-1} &= -\sum_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}p}\{(1-p)^x\} = -\frac{\mathrm{d}}{\mathrm{d}p}\sum_{x=0}^{\infty}(1-p)^x \\ &= -\frac{\mathrm{d}}{\mathrm{d}p}\sum_{x=0}^{\infty}(1-p)^x = -\frac{\mathrm{d}}{\mathrm{d}p}\frac{1}{1-(1-p)} \\ &= -\frac{\mathrm{d}}{\mathrm{d}p}\left\{\frac{1}{p}\right\} = -\left(-\frac{1}{p^2}\right) = \frac{1}{p^2}\end{aligned}$$

y

$$\mathrm{E}\,X = p\frac{1}{p^2} = \frac{1}{p}.$$

Al considerar (2.29), la varianza se obtiene de manera similar:

$$\begin{aligned}\mathrm{E}\,X^2 &= \sum_{x=0}^{\infty} x^2 f(x) = \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} \\ &= p\sum_{x=1}^{\infty} x(x-1+1)(1-p)^{x-1} \\ &= p\sum_{x=1}^{\infty} x(x-1)(1-p)^{x-1} + p\sum_{x=1}^{\infty} x(1-p)^{x-1} \\ &= p\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-2+1} + \mathrm{E}\,X \\ &= p(1-p)\sum_{x=0}^{\infty} x(x-1)(1-p)^{x-2} + \frac{1}{p}.\end{aligned}$$

Los términos de la anterior sumatoria tienen forma de la segunda derivada de una función polinomial, incluso desde el índice $x=0$:

$$\frac{\mathrm{d}^2}{\mathrm{d}\,p^2}\{(1-p)^x\} = \frac{\mathrm{d}}{\mathrm{d}p}\{-x(1-p)^{x-1}\} = x(x-1)(1-p)^{x-2}; \quad x = 0, 1, \ldots$$

Así

$$\begin{aligned}\sum_{x=0}^{\infty} x(x-1)(1-p)^{x-2} &= \sum_{x=0}^{\infty}\frac{\mathrm{d}^2}{\mathrm{d}\,p^2}\{(1-p)^x\} = \frac{\mathrm{d}^2}{\mathrm{d}\,p^2}\sum_{x=0}^{\infty}(1-p)^x \\ &= \frac{\mathrm{d}^2}{\mathrm{d}\,p^2}\frac{1}{1-(1-p)} = \frac{\mathrm{d}^2}{\mathrm{d}\,p^2}\left\{\frac{1}{p}\right\} \\ &= \frac{\mathrm{d}}{\mathrm{d}p}\left\{-\frac{1}{p^2}\right\} = \frac{2}{p^3},\end{aligned}$$

$$\begin{aligned}\mathrm{E}\,X^2 &= p(1-p)\frac{2}{p^3} + \frac{1}{p} = \frac{2(1-p)}{p^2} + \frac{1}{p} \\ &= \frac{2(1-p)+p}{p^2} = \frac{2-p}{p^2}\end{aligned}$$

y

$$
\begin{aligned}
\mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 &= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\
&= \frac{(2-p)-1}{p^2} = \frac{1-p}{p^2}.
\end{aligned}
$$

Ahora bien, si Y representa el número de fracasos hasta el primer éxito, entonces $Y = X - 1$. sus valores posibles son $Y = 0, 1, 2, \ldots$ Además

$$
\begin{aligned}
\mathrm{E}\, Y &= \mathrm{E}[X-1] = \mathrm{E}\, X - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}, \\
\mathrm{V}\, Y &= \mathrm{V}[X-1] = \mathrm{V}\, X = \frac{1-p}{p^2}.
\end{aligned}
$$

Note que la esperanza de ambas variables aleatorias está desfasada por una unidad. En cambio, sus varianzas son iguales.

Por ejemplo, sea X el número de boletos comprados hasta ganar el premio mayor de la Lotería Nacional. La probabilidad de ganar en cada sorteo es $p = 1/50\,000 = 0.000\,02$. El número de intentos esperados es

$$
\mathrm{E}\, X = \frac{1}{p} = \frac{1}{1/50000} = 50\,000 \text{ juegos.}
$$

La varianza es

$$
\mathrm{V}\, X = \frac{1-p}{p^2} = \frac{1-1/50000}{(1/50000)^2} = 2.499\,95 \times 10^9 \text{ juegos}^2.
$$

La desviación estándar es

$$
\mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\, X} = \sqrt{2.49995 \times 10^9} = 49\,999.5 \text{ juegos.}
$$

En otro orden de ideas, la *mediana* o *mediana poblacional*, es un valor $m \in \mathbf{R}$ que satisface

$$
P(X \leq m) \geq \frac{1}{2} \quad \text{y} \quad P(X \geq m) \geq \frac{1}{2}. \tag{2.36}
$$

Como la media poblacional, la mediana es una medida de tendencia central, aunque puede no ser única, especialmente para poblaciones discretas. Por ejemplo, si $X \sim \mathrm{U}(1, \ldots, 10)$, entonces la mediana es cualquier número real m del intervalo semicerrado $[5, 6)$:

$$
P(X \leq m) = P(X \leq 5) = \frac{5}{10} = 0.5
$$

y

$$
P(X \geq m) \geq P(X \geq 6) = 1 - P(X < 6) = 1 - P(X \leq 5) = 0.5.
$$

Una alternativa de definición única de *mediana* es

$$
m = \operatorname{ínf}\left\{x \in \mathbf{R} : P(X \leq x) \geq \frac{1}{2}\right\} = \sup\left\{x \in \mathbf{R} : P(X \leq x) < \frac{1}{2}\right\}. \tag{2.37}
$$

En ese sentido, $m = 5$ es la mediana de la distribución uniforme discreta $\mathrm{U}(1, \ldots, 10)$.

Otra medida de tendencia central es la moda. Para una población discreta, la *moda* es aquel valor posible con la mayor probabilidad:

$$moda = \arg\max\{f(x) : x \in \mathbf{R}\}. \tag{2.38}$$

Así

$$P(X = moda) \geq P(X = x); \quad \text{para todo} \quad x \in \mathbf{R}.$$

La moda puede no ser única. Para el ejemplo de la distribución uniforme discreta $\mathrm{U}(1, \ldots, 10)$, cada uno de sus 10 valores posibles es moda.

2.4. Vectores aleatorios discretos

Esta sección es una introducción a las vectores de dos o más variables aleatorias. Por el contexto del capítulo, se hace énfasis al caso discreto.

Como introducción, se definirá a un vector aleatorio de dimensión dos, así como sus correspondientes funciones de densidad marginales y conjunta. Se abordará también la medida de dependencia lineal llamada correlación.

Una población de interés se explica en términos de una, dos o más variables aleatorias, según el número de dimensiones con que se observa el fenómeno de estudio. En este sentido, sean X y Y dos variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces, (X, Y) es un *vector aleatorio* de dimensión dos. Como en el caso univariado, un vector aleatorio (X, Y) tiene asociada una función de *distribución* o *distribución conjunta*:

$$F(x, y) = P(X \leq x, Y \leq y); \quad (x, y) \in \mathbf{R}^2. \tag{2.39}$$

Análogo al Teorema 2.1.3 del caso unidimensional, la función $F(x, y)$ se caracteriza por las siguientes tres propiedades:

1. Si $x_1 \leq x_2$ y $y_1 \leq y_2$, entonces

$$F(x_1, x_2) \leq F(y_1, y_2).$$

2.

$$F(-\infty, -\infty) = \lim_{x,y \to -\infty} F(x, y) = 0$$

y

$$F(\infty, \infty) = \lim_{x,y \to \infty} F(x, y) = 1.$$

3. Si las sucesiones $\{x_n\}$, $\{y_n\}$; $n \geq 1$, decrecen respectivamente a $x, y \in \mathbf{R}$, cuando $n \to \infty$, entonces

$$F(x, y) = \lim_{n \to \infty} F(x_n, y_n).$$

La última propiedad se debe a la continuidad de la probabilidad (1.24). Del mismo modo, por (1.23), aplica un resultado análogo a (2.9). Si las sucesiones $\{x_n\}$, $\{y_n\}$; $n \geq 1$, crecen respectivamente a $x, y \in \mathbf{R}$, cuando $n \to \infty$, entonces

$$P(X < x, Y < y) = \lim_{n\to\infty} F(x_n, y_n) \leq F(x, y).$$

Ahora se desarrollarán las propiedades de los vectores aleatorios discretos. Se dice que (X, Y) es un *vector aleatorio discreto* si ambas variables aleatorias son discretas. Salvo que se mencione lo contrario, se considera que el conjunto de *vectores posibles* de (X, Y) es $(\mathbf{N}\cup\{0\})^2$:

$$X, Y = 0, 1, 2, \dots$$

La Figura 2.10[izquierda] representa a dicho conjunto de vectores posibles. La función de *densidad* o *densidad conjunta* de (X, Y) es

$$f(x, y) = P(X = x, Y = y); \quad x, y = 0, 1, 2, \dots \tag{2.40}$$

Tanto la función de distribución conjunta (2.39) como su función de densidad conjunta (2.40) determinan la distribución del vector aleatorio discreto (X, Y). Por ejemplo, de la ley de probabilidad total (1.19), se tiene

$$P((X, Y) \in A) = \sum_{(x,y)\in A} f(x, y); \quad A \in \mathcal{B}(\mathbf{R}^2).$$

Aquí A es cualquier región o conjunto de Borel en el plano. En particular

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{i=0}^{x}\sum_{j=0}^{y} P(X = i, Y = j) = \sum_{i=0}^{x}\sum_{j=0}^{y} f(i, j);$$

para $x, y = 0, 1, 2, \dots$ Si el evento de interés es $\{X \leq Y\}$, entonces

$$A = \{(x, y) \in \mathbf{Z}^2 : 0 \leq x \leq y\}$$

y

$$P(X \leq Y) = \sum_{(x,y)\in A} f(x, y) = \sum_{x=0}^{\infty}\sum_{y=x}^{\infty} f(x, y) = \sum_{y=0}^{\infty}\sum_{x=0}^{y} f(x, y).$$

El intercambio de los índices de las sumas se debe al teorema de Tonelli Fubini, pues sus términos son no negativos; véase [1, Teorema 8.43]. La Figura 2.10[izquierda] describe la región en el plano del evento $\{X \leq Y\}$.

Análogo a las propiedades del Teorema 2.1.6, para variables aleatorias discretas, una función de densidad conjunta $f(x, y)$, satisface las siguientes dos propiedades características:

$$f(x, y) \geq 0; \quad \text{para todo} \quad x, y = 0, 1, \dots \tag{2.41}$$

y

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x, y) = \sum_{y=0}^{\infty}\sum_{x=0}^{\infty} f(x, y) = 1. \tag{2.42}$$

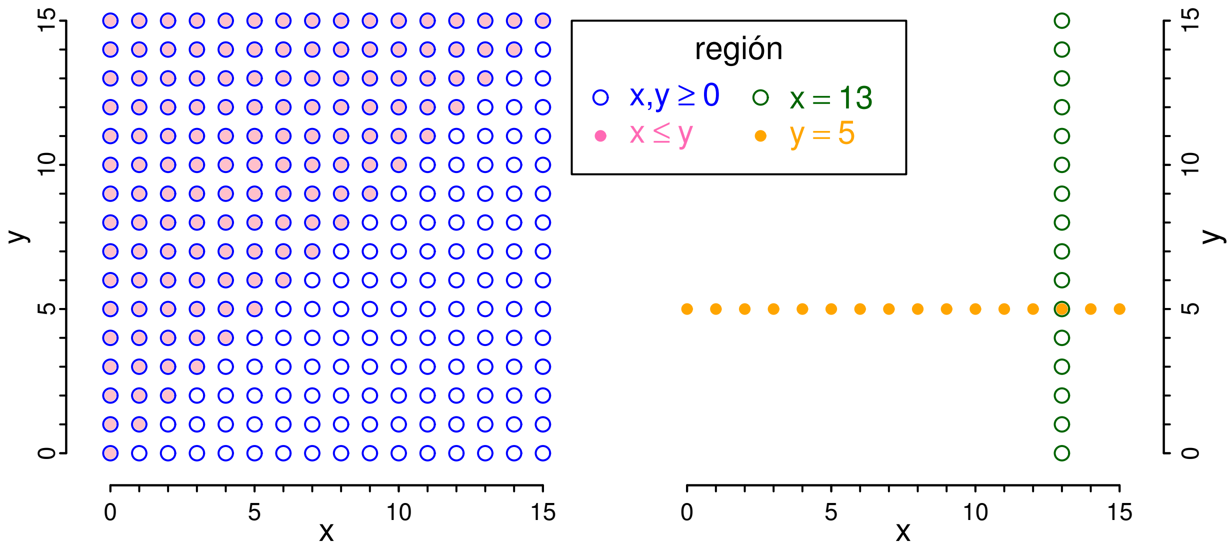


Figura 2.10. Representación gráfica del conjunto de vectores posibles de (X, Y), así como de los evento $\{X \leq Y\}$, $\{X = x\}$, $\{Y = y\}$, con $x = 13$ y $y = 5$.

En un contexto bivariado, a las funciones de densidad de X y de Y, se les dice funciones de densidad marginales. La función de *densidad marginal* de la variable aleatoria X es

$$f(x) = f_X(x) = P(X = x); \quad x = 0, 1, 2, \ldots$$

Así mismo, la función de *densidad marginal* de la variable aleatoria Y es

$$f(y) = f_Y(y) = P(Y = y); \quad y = 0, 1, 2, \ldots$$

La notación con subíndice evita ambigüedades: $f_X(x)$, $f_Y(y)$, $f_{X,Y}(x, y)$. Sin embargo, es usual la omisión de tal subíndice con el entendido de que el argumento con letras minúsculas especifica la función de densidad de la variable o vector aleatorio de letras mayúsculas correspondientes. De la ley de probabilidad total (1.19), ambas densidades marginales se deducen por la función de densidad conjunta:

$$f(x) = \sum_{y=0}^{\infty} f(x, y); \quad x = 0, 1, 2, \ldots, \tag{2.43}$$

$$f(y) = \sum_{x=0}^{\infty} f(x, y); \quad y = 0, 1, 2, \ldots \tag{2.44}$$

De hecho, para $x \geq 0$, el evento $\{X = x\}$ es una unión de eventos excluyentes:

$$\begin{aligned}
\{X = x\} &= \{X = x\} \cap \Omega = \{X = x\} \cap \{Y \in \mathbf{R}\} \\
&= \{X = x\} \cap \left(\bigcup_{y=0}^{\infty} \{Y = y\} \right) = \bigcup_{y=0}^{\infty} (\{X = x\} \cap \{Y = y\}) \\
&= \bigcup_{y=0}^{\infty} \{X = x, Y = y\}.
\end{aligned}$$

Por lo que

$$f(x) = P(X = x) = \sum_{y=0}^{\infty} P(X = x, Y = y) = \sum_{y=0}^{\infty} f(x, y); \quad x = 0, 1, 2, \ldots$$

Aquí se barre sobre todos los enteros $y \geq 0$. En la Figura 2.10[derecha] se aprecia la gráfica del evento $\{X = x\}$, con $x = 13$; punto hueco verde. Una deducción similar aplica para la función de densidad marginal de Y (2.44). En dicha figura se nuestra también la gráfica del evento $\{Y = y\}$, con $y = 5$; punto naranja.

A continuación se desarrolla una extensión de la fórmula (2.26). Sea $g(x, y)$ una función real definida para $x, y = 0, 1, 2, \ldots$, tal que

$$\sum_{x=0}^{\infty} \sum_{y=0}^{\infty} |g(x, y)| f(x, y) < \infty.$$

Entonces

$$\mathrm{E}\, g(X, Y) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} g(x, y) f(x, y) = \sum_{y=0}^{\infty} \sum_{x=0}^{\infty} g(x, y) f(x, y). \tag{2.45}$$

De nuevo, el intercambio de los índices de las sumas se debe al teorema de Tonelli Fubini; véase [1, Teorema 8.43]. En consecuencia, las esperanzas de X y Y se representan en términos tanto de la función de densidad conjunta como de sus respectivas densidades marginales:

$$\mathrm{E}\, X = \sum_{x=0}^{\infty} x f(x) = \sum_{x=0}^{\infty} x \sum_{y=0}^{\infty} f(x, y) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} x f(x, y) = \sum_{y=0}^{\infty} \sum_{x=0}^{\infty} x f(x, y)$$

y

$$\mathrm{E}\, Y = \sum_{y=0}^{\infty} y f(y) = \sum_{y=0}^{\infty} y \sum_{x=0}^{\infty} f(x, y) = \sum_{y=0}^{\infty} \sum_{x=0}^{\infty} y f(x, y) = \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} y f(x, y).$$

Aquí se aplicó (2.43)-(2.45).

Ejemplo 2.4.1. Considere una variable aleatoria de distribución binomial $X \sim \mathrm{bin}(n, p)$, donde $n \geq 0$ es el número de éxitos y $0 \leq p \leq 1$ es la probabilidad de éxito en cada ensayo de Bernoulli. El total de fracasos $Y = n - X$, tiene también distribución binomial, de parámetros n y $1 - p$. Ambas variables aleatorias comparten el conjunto de valores posibles:

$$X, Y = 0, 1, \ldots, n.$$

Sin embargo, conjuntamente los vectores posibles de (X, Y) son

$$(X, Y) = (0, n), (1, n - 1), \ldots, (n - 1, 1), (n, 0).$$

Este conjunto se aprecia en la Figura 2.11, para $n = 5$ ensayos; punto rojo. Si $X = x$, entonces $Y = n - x$. Viceversa, $X = n - y$, siempre que $Y = y$. La función de densidad conjunta de (X, Y) es

$$f(x, y) = P(X = x, Y = y) = \binom{n}{x} p^x (1 - p)^y; \quad \text{para} \quad x = 0, \ldots, n, \quad y = n - x.$$

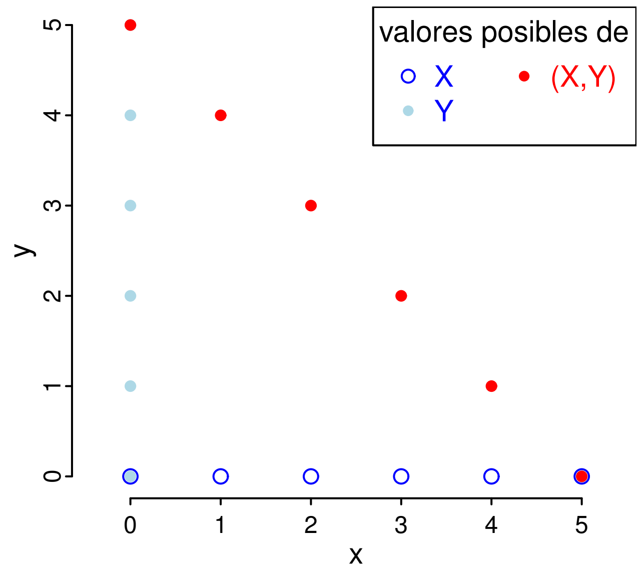


Figura 2.11. Gráfica del conjunto de vectores posibles de (X, Y), junto con los valores posibles de las variables aleatorias discretas X y $Y = n - X$, con $n = 5$; Ejemplo 2.4.1.

En otros casos, la densidad es cero. Se confirma que las funciones de densidad marginales son binomiales: distribución discreta!binomial

$$f(x) = \sum_{y=0}^{\infty} f(x, y) = f(x, n-x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, \ldots, n$$

y

$$\begin{aligned} f(y) &= \sum_{x=0}^{\infty} f(x, y) = f(n-y, y) \\ &= \binom{n}{n-y} p^{n-y}(1-p)^{n-(n-y)} = \binom{n}{y} p^{n-y}(1-p)^y \\ &= \binom{n}{y} (1-p)^y [1-(1-p)]^{n-y}; \quad y = 0, \ldots, n. \end{aligned}$$

Por lo tanto, $X \sim \text{bin}(n, p)$ y $Y \sim \text{bin}(n, 1-p)$. Por la relación lineal que guardan las variables aleatorias X y Y, su *covarianza* es negativa mientras que el coeficiente de *correlación* es -1; véase Ejercicio 2.11.9.

Sean X y Y dos variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, no necesariamente discretas, con funciones de distribución conjunta y marginales $F(x, y) = P(X \leq x, Y \leq y)$, $F(x) = P(X \leq x)$, $F(y) = P(Y \leq y)$; para $x, y \in \mathbf{R}$. Se dice que las variables aleatorias X y Y son *independientes* si

$$F(x, y) = F(x)F(y); \quad x, y \in \mathbf{R}. \tag{2.46}$$

El siguiente teorema caracteriza la independencia de dos variables aleatorias.

Teorema 2.4.2. Sean X y Y dos variables aleatorias. Entonces, las siguientes afirmaciones son equivalentes:

1. X y Y son variables aleatorias independientes

2. $$F(x, y) = F(x)F(y); \quad x, y \in \mathbf{R},$$

3. $$P(X < x, Y < y) = P(X < x)P(Y < y); \quad x, y \in \mathbf{R},$$

4. $$P(X > x, Y > y) = P(X > x)P(Y > y); \quad x, y \in \mathbf{R},$$

5. $$P(X \geq x, Y \geq y) = P(X \geq x)P(Y \geq y); \quad x, y \in \mathbf{R},$$

6. $$P(X \in A, Y \in B) = P(X \in A)P(Y \in B); \quad A, B \in \mathcal{B}(\mathbf{R}).$$

Para el caso discreto, la independencia de dos variables aleatorias se caracteriza también por la igualdad de la función de densidad conjunta con el producto de sus funciones de densidad marginales, como se afirma en el siguiente corolario. Se definirá antes el concepto de densidad condicional.

Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta $f(x, y)$ (2.40). La función de *densidad condicional* de Y dado el evento $\{X = x\}$ es

$$f(y \mid x) = P(Y = y \mid X = x) = \frac{f(x, y)}{f(x)}; \quad x, y = 0, 1, \ldots,$$

siempre que $f(x) > 0$, y cero en otros casos. Por (1.43), recuerde que

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{f(x)}.$$

De la misma manera, dado el evento $\{Y = y\}$, la función de *densidad condicional* de la variable aleatoria X es

$$f(x \mid y) = P(X = x \mid Y = y) = \frac{f(x, y)}{f(y)}; \quad x, y = 0, 1, \ldots,$$

siempre que $f(y) > 0$, y cero en otros casos. Así, la función de densidad conjunta de X y Y se escribe de dos maneras diferentes:

$$f(x, y) = f(x)f(y \mid x) = f(x \mid y)f(y); \quad \text{para} \quad x, y = 0, 1, \ldots$$

Corolario 2.4.3. Sean X y Y dos variables aleatorias discretas, con conjunto de valores posibles en los enteros no negativos. Entonces, las afirmaciones del Teorema 2.4.2 son equivalentes sus contraparte evaluadas en $x, y = 0, 1, \ldots,$ así como $A, B \subset \{0, 1, \ldots\}$. Además, dichas afirmaciones son equivalentes a las siguientes.

1.
$$f(x,y) = f(x)f(y); \quad x,y = 0,1,2,\dots \tag{2.47}$$

2.
$$f(y \mid x) = f(y); \quad x,y = 0,1,\dots \quad \text{con} \quad f(x) > 0,$$

3.
$$f(x \mid y) = f(x); \quad x,y = 0,1,\dots, \quad \text{con} \quad f(y) > 0,$$

Demostración. Por el Ejercicio 2.11.6, se demuestra la equivalencia de la independencia de las variables aleatorias discretas X y Y con (2.46). Además, X y Y son variables aleatorias independientes si y sólo si

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{f(x)f(y)}{f(x)} = f(y) \quad x,y = 0,1,\dots$$

La otra equivalencia se verifica de manera similar. □

El siguiente teorema resume las propiedades básicas del operador esperanza.

Teorema 2.4.4. Sean X y Y dos variables aleatorias con esperanza finita (no necesariamente discretas). Considere las constantes reales a y b. Entonces

1.
$$\mathrm{E}\, a = a.$$

2. El operador esperanza es una transformación lineal:
$$\mathrm{E}[aX] = a\,\mathrm{E}\,X \quad \text{y} \quad \mathrm{E}[X+Y] = \mathrm{E}\,X + \mathrm{E}\,Y. \tag{2.48}$$
Esta propiedad equivale a cualquiera de las siguientes dos afirmaciones
$$\mathrm{E}[aX + bY] = a\,\mathrm{E}\,X + b\,\mathrm{E}\,Y$$
y
$$\mathrm{E}[a_1X_1 + \cdots + a_nX_n] = a_1\,\mathrm{E}\,X_1 + \cdots + a_n\,\mathrm{E}\,X_n.$$
donde $X_1,\dots,X_n$ son $n \geq 1$ variables aleatorias con esperanza finita y $a_1,\dots,a_n$ son constantes. En particular
$$\mathrm{E}[X_1 + \cdots + X_n] = \mathrm{E}\,X_1 + \cdots + \mathrm{E}\,X_n. \tag{2.49}$$

3. Si X y Y son variables aleatorias independientes, entonces
$$\mathrm{E}[XY] = \mathrm{E}\,X\,\mathrm{E}\,Y. \tag{2.50}$$
El recíproco no es cierto.

4. Si $X \leq Y$, es decir $P(X \leq Y) = 1$, entonces

$$\mathrm{E}\,X \leq \mathrm{E}\,Y.$$

El recíproco no es cierto. En particular

a) Si $X \geq 0$, entonces $\mathrm{E}\,X \geq 0$. Además

$$\mathrm{E}\,X = 0, \quad \text{si y sólo si} \quad X = 0.$$

b) Si $a \leq X \leq b$, entonces

$$a \leq \mathrm{E}\,X \leq b.$$

c)

$$|\,\mathrm{E}\,X| \leq \mathrm{E}\,|X|.$$

Demostración. Algunas propiedades se verificarán sólo para el caso de variables aleatorias discretas. El caso continuo se aborda en el Capítulo 3.
1. Si la variable aleatoria X es la constante a, entonces su función de densidad concentra su *masa* en $x = a$:

$$P(X = a) = 1.$$

Así

$$\mathrm{E}\,X = a \cdot P(x = a) = a \cdot 1 = a.$$

4. Las propiedades 4., 4*b*. y 4*c*. se deducen por 4*a*.; véase Ejercicio 2.7. Esta última propiedad se verifica como sigue. Si $P(X \geq 0) = 1$, entonces

$$\mathrm{E}\,X = \sum_{x=0}^{\infty} x f(x) \geq 0.$$

Si además $\mathrm{E}\,X = 0$, entonces

$$\mathrm{E}\,X = \sum_{x=0}^{\infty} x f(x) = 0f(0) + 1f(1) + 2f(2) + \cdots = 0,$$

con $f(0), f(1), f(2), \cdots \geq 0$. Por lo que

$$f(1) = f(2) = f(3) = \cdots = 0.$$

Por otro lado, recuerde

$$f(0) + f(1) + f(2) + \cdots = 1.$$

Así, $f(0) = 1$ y X es la constante cero.
2. Primero se comprobará $\mathrm{E}\,|aX| < \infty$. Como $\mathrm{E}\,|X| < \infty$, entonces

$$\mathrm{E}\,|aX| = \sum_x |ax| f(x) = \sum_x |a||x| f(x) = |a| \sum_x |x| f(x) = |a|\,\mathrm{E}\,|X| < \infty.$$

Así que

$$\mathrm{E}(aX) = \sum_x axf(x) = a\sum_x xf(x) = a\,\mathrm{E}\,X.$$

Para comprobar la segunda fórmula de (2.48), considere la función de densidad conjunta (2.40) $f(x,y)$. La desigualdad del triángulo se satisface para cada evento simple $\omega \in \Omega$:

$$|X(\omega) + Y(\omega)| \le |X(\omega)| + |Y(\omega)|,$$

donde $|X|$ y $|Y|$ son variables aleatorias de esperanza finita. Por la propiedad 4, se tiene

$$\begin{aligned}\mathrm{E}\,|X+Y| \le \mathrm{E}(|X| + \mathrm{E}\,|Y|) &= \sum_x\sum_y(|x|+|y|)f(x,y)\\ &= \sum_x\sum_y |x|f(x,y) + \sum_x\sum_y |y|f(x,y) = \mathrm{E}\,|X| + \mathrm{E}\,|Y|\\ &< \infty.\end{aligned}$$

Por lo tanto

$$\begin{aligned}\mathrm{E}(X+Y) &= \sum_x\sum_y(x+y)f(x,y) = \sum_x\sum_y xf(x,y) + \sum_x\sum_y yf(x,y)\\ &= \mathrm{E}\,X + \mathrm{E}\,Y.\end{aligned}$$

La expresión (2.49) se verifica por inducción.
3. Por la hipótesis de independencia, la función de densidad conjunta es el producto de sus funciones de densidad marginales (2.47). Así

$$\begin{aligned}\mathrm{E}\,|XY| &= \sum_x\sum_y |xy|f(x,y) = \sum_x\sum_y |x||y|f(x)f(y)\\ &= \sum_x |x|f(x)\sum_y |y|f(y) = \mathrm{E}\,|X|\,\mathrm{E}\,|Y| < \infty\end{aligned}$$

y

$$\begin{aligned}\mathrm{E}[XY] &= \sum_x\sum_y xyf(x,y) = \sum_x\sum_y xyf(x)f(y)\\ &= \sum_x xf(x)\sum_y yf(y) = \mathrm{E}\,X\,\mathrm{E}\,Y.\end{aligned}$$

□

Como consecuencia de este teorema, se verifica la fórmula simplificada de la varianza (2.29):

$$\begin{aligned}\mathrm{E}(X-\mu)^2 &= \mathrm{E}[X^2 - 2\mu X + \mu^2]\\ &= \mathrm{E}\,X^2 + \mathrm{E}[-2\mu X] + \mathrm{E}\,\mu^2\\ &= \mathrm{E}\,X^2 - 2\mu\,\mathrm{E}\,X + \mu^2\\ &= \mathrm{E}\,X^2 - 2\mu\mu + \mu^2\\ &= \mathrm{E}\,X^2 - \mu^2.\end{aligned}$$

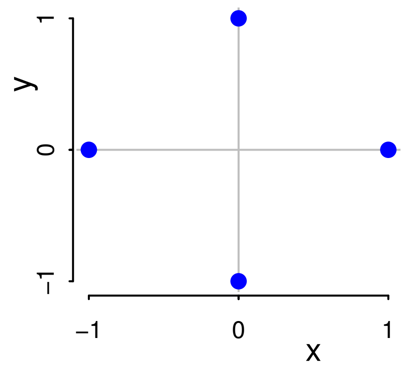


Figura 2.12. Conjunto de vectores posibles de un vector aleatorio (X,Y), cuyos elementos son variables aleatorias dependientes y con cero covarianza; Ejemplo 2.4.5.

Ejemplo 2.4.5. Por (2.50) del Teorema 2.4.4, se tiene $\mathrm{E}[XY] = \mathrm{E}\,X\,\mathrm{E}\,Y$, siempre que X y Y sean variables aleatorias independientes. Aquí se muestra un contraejemplo de que el recíproco no es cierto. Suponga que el vector aleatorio (X,Y) tiene distribución uniforme discreta en los cuatro puntos $(-1,0)$, $(0,-1)$, $(1,0)$ y $(0,1)$. En la Figura 2.12 se muestra la gráfica de los vectores posibles de (X,Y). Con probabilidad uno, se tiene $XY=0$. Además

$$\mathrm{E}\,X = \mathrm{E}\,Y = 0.$$

Entonces

$$\mathrm{E}[XY] = 0 = \mathrm{E}\,X\,\mathrm{E}\,Y.$$

Sin embargo, X y Y no son variables aleatorias independientes, ya que

$$P(X=0,Y=0) = 0 < \frac{1}{2}\cdot\frac{1}{2} = P(X=0)P(Y=0).$$

2.5. Covarianza y correlación

En esta sección se definen los coeficientes de covarianza y correlación. De estos conceptos, junto con la varianza, se muestran sus propiedades básicas.

Si X y Y son dos variables aleatorias de varianza finita, la *covarianza* de X y Y es

$$\mathrm{C}(X,Y) = \mathrm{E}[(X-\mathrm{E}\,X)(Y-\mathrm{E}\,Y)]. \tag{2.51}$$

Si además $\mathrm{V}\,X, \mathrm{V}\,Y > 0$, el coeficiente de *correlación* es

$$\rho(X,Y) = \frac{\mathrm{C}(X,Y)}{\sqrt{\mathrm{V}\,X\cdot\mathrm{V}\,Y}}. \tag{2.52}$$

Este coeficiente mide la dependencia lineal entre las variables aleatorias X y Y. Si X o Y es una constante, entonces la correlación no está definida.

En la siguiente proposición se describen algunas propiedades de la varianza, covarianza y el coeficiente de correlación. Para este último, se asume que las varianzas involucradas son positivas. Para mayores propiedades de la covarianza y la correlación véase la Sección 4, más adelante.

Teorema 2.5.1. Considere que las variables aleatorias invocadas, no necesariamente discretas, tienen varianza finita. Sean las constantes reales $a, a_1, \ldots, a_m$ y $b, b_1, \ldots, b_n$, con $m, n \geq 1$. Entonces

1. $$\mathrm{V}\, X \geq 0.$$
   En particular, $\mathrm{V}\, X = 0$, si y sólo si X es una constante.

2. $$\mathrm{V}[X+Y] = \mathrm{V}\, X + \mathrm{V}\, Y + 2\, \mathrm{C}(X,Y)$$
   y
   $$\mathrm{V}[X_1 + \cdots + X_n] = \sum_{i=1}^{n} \mathrm{V}\, X_i + 2 \sum_{1 \leq i < j \leq n} \mathrm{C}(X_i, X_j).$$

3. $$\mathrm{C}(X,Y) = \mathrm{E}[XY] - \mathrm{E}\, X\, \mathrm{E}\, Y. \tag{2.53}$$
   En particular
   $$\mathrm{C}(X,X) = \mathrm{V}\, X. \tag{2.54}$$

4. La covarianza es simétrica
   $$\mathrm{C}(X,Y) = \mathrm{C}(Y,X). \tag{2.55}$$

5. La covarianza es bilineal
   $$\begin{aligned} \mathrm{C}(aX_1 + X_2, Y) &= a\, \mathrm{C}(X_1, Y) + \mathrm{C}(X_2, Y), \\ \mathrm{C}(X, bY_1 + Y_2) &= b\, \mathrm{C}(X, Y_1) + \mathrm{C}(X, Y_2). \end{aligned}$$
   Esta propiedad equivale a cualquiera de las siguientes dos afirmaciones
   $$\begin{aligned} &\mathrm{C}(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) \\ &= a_1b_1\, \mathrm{C}(X_1,Y_1) + a_1b_2\, \mathrm{C}(X_1,Y_2) + a_2b_1\, \mathrm{C}(X_2,Y_1) + a_2b_2\, \mathrm{C}(X_2,Y_2), \end{aligned} \tag{2.56}$$
   $$\mathrm{C}\left(\sum_{i=1}^{m} a_iX_i, \sum_{j=1}^{n} b_jY_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j\, \mathrm{C}(X_i,Y_j), \tag{2.57}$$
   donde $X_1, \ldots, X_m$ y $Y_1, \ldots, Y_n$ son variables aleatorias de varianza finita; para $m, n \geq 1$.

6. Si X y Y son variables aleatorias independientes, entonces
$$\mathrm{C}(X,Y)=0$$
y
$$\mathrm{V}[X+Y]=\mathrm{V}\,X+\mathrm{V}\,Y.$$
El recíproco no es cierto. En particular
$$\mathrm{C}(X,a)=0$$
y
$$\mathrm{V}[a+bX]=b^2\,\mathrm{V}\,X.$$
7. La correlación es simétrica
$$\rho(X,Y)=\rho(Y,X).$$
8. $$\rho(a_1X+a_2,b_1Y+b_2)=\mathrm{signo}(a_1b_1)\rho(X,Y).$$
En particular, el coeficiente de correlación no depende de la escala de las variables aleatorias:
$$\rho(X,Y)=\rho\left(\frac{X-\mu_1}{\sigma_1},\frac{Y-\mu_2}{\sigma_2}\right)=\mathrm{C}\left(\frac{X-\mu_1}{\sigma_1},\frac{Y-\mu_2}{\sigma_2}\right), \tag{2.58}$$
donde $\mu_1=\mathrm{E}\,X$, $\mu_2=\mathrm{E}\,Y$, $\sigma_1^2=\mathrm{V}\,X$ y $\sigma_2^2=\mathrm{V}\,Y$.
9. $$|\,\mathrm{C}(X,Y)|\leq\sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y}.$$
Si además $\mathrm{V}\,X,\mathrm{V}\,Y>0$, esta desigualdad equivale a
$$-1\leq\rho(X,Y)\leq 1.$$
10. Si $b\neq 0$, entonces
$$Y=a+bX\quad\Leftrightarrow\quad\rho(X,Y)=\mathrm{signo}\,b=\pm 1.$$
En particular
$$\rho(X,X)=1\quad\text{y}\quad\rho(X,-X)=-1.$$

Demostración. Algunas propiedades se verificarán sólo para el caso de variables aleatorias discretas. El caso continuo se aborda en el Capítulo 3.
1. Por definición de la varianza (2.28), esta es no negativa:
$$\mathrm{V}\,X=\mathrm{E}(X-\mu)^2\geq 0.$$
Por el Teorema 2.4.4.4a, se tienen las siguientes equivalencias:
$$\begin{aligned}\mathrm{E}(X-\mu)^2&=0,\\(X-\mu)^2&=0,\\X-\mu&=0,\\X&=\mu.\end{aligned}$$

4. La propiedad de simetría de la covarianza (2.55) es inmediata.
5. Por tal propiedad de simetría, es suficiente verificar la bilinealidad (2.56) con la linealidad de la covarianza en cualquiera de sus dos coordenadas. Así

$$\begin{aligned}
\mathrm{C}(aX_1 + X_2, Y) &= \mathrm{E}[(aX_1 + X_2)Y] - \mathrm{E}(aX_1 + X_2)\,\mathrm{E}\,Y \\
&= \mathrm{E}[aX_1Y + X_2Y] - (a\,\mathrm{E}\,X_1 + \mathrm{E}\,X_2)\,\mathrm{E}\,Y \\
&= a\,\mathrm{E}[X_1Y] + \mathrm{E}[X_2Y] - a\,\mathrm{E}\,X_1\,\mathrm{E}\,Y - \mathrm{E}\,X_2\,\mathrm{E}\,Y \\
&= a(\mathrm{E}[X_1Y] - \mathrm{E}\,X_1\,\mathrm{E}\,Y) + (\mathrm{E}[X_2Y] - \mathrm{E}\,X_2\,\mathrm{E}\,Y) \\
&= a\,\mathrm{C}(X_1, Y) + \mathrm{C}(X_2, Y).
\end{aligned}$$

2. Este resultado se deduce por las propiedades (2.54), simetría (2.55) y bilinealidad (2.56):

$$\begin{aligned}
\mathrm{V}(X + Y) &= C(X + Y, X + Y) \\
&= C(X, X) + C(X, Y) + C(Y, X) + C(Y, Y) \\
&= \mathrm{V}\,X + \mathrm{V}\,Y + 2\,\mathrm{C}(X, Y)
\end{aligned}$$

Así mismo, por (2.57), resulta

$$\begin{aligned}
\mathrm{V}[X_1 + \cdots + X_n] &= \mathrm{C}(X_1 + \cdots + X_n, X_1 + \cdots + X_n) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{C}(X_i, X_j) \\
&= \sum_{i=1}^{n} \mathrm{V}\,X_i + 2 \sum_{1 \le i < j \le n} \mathrm{C}(X_i, X_j).
\end{aligned}$$

3. Si $\mathrm{E}\,X = \mu_1$ y $\mathrm{E}\,Y = \mu_2$, entonces

$$\begin{aligned}
\mathrm{C}(X, Y) &= \mathrm{E}[(X - \mu_1)(Y - \mu_2)] \\
&= \mathrm{E}[XY - \mu_1 Y - X\mu_2 + \mu_1\mu_2] \\
&= \mathrm{E}\,XY - \mu_1\,\mathrm{E}\,Y - \mu_2\,\mathrm{E}\,X + \mu_1\mu_2 \\
&= \mathrm{E}\,XY - \mu_1\mu_2.
\end{aligned}$$

En particular

$$\mathrm{C}(X, X) = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = \mathrm{V}\,X.$$

6. Si X y Y son variables aleatorias independientes, entonces

$$\mathrm{C}(X, Y) = \mathrm{E}[XY] - \mathrm{E}\,X\,\mathrm{E}\,Y = \mathrm{E}\,X\,\mathrm{E}\,Y - \mathrm{E}\,X\,\mathrm{E}\,Y = 0.$$

En particular, si c es una constante, entonces X y c son variables aleatorias independientes. Así

$$\mathrm{C}(X, c) = 0 \quad \text{y} \quad \mathrm{V}(a + bX) = b^2\,\mathrm{V}\,X.$$

Por otro lado, del Ejemplo 2.4.5, se verifica que el recíproco no es cierto.
7. La propiedad de simetría de la covarianza se transmite a la correlación.

8. Considere

$$\begin{aligned}
\rho(a_1X + a_2, b_1Y + b_2) &= \frac{\mathrm{C}(a_1X + a_2, b_1Y + b_2)}{\sqrt{\mathrm{V}(a_1X + a_2)\,\mathrm{V}(b_1Y + b_2)}} \\
&= \frac{\mathrm{C}(a_1X + a_2, b_1Y) + \mathrm{C}(a_1X + a_2, b_2)}{\sqrt{a_1^2\,\mathrm{V}\,X \cdot b_1^2\,\mathrm{V}\,Y}} \\
&= \frac{a_1b_1\,\mathrm{C}(X, Y) + 0}{|a_1b_1|\sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y}} \\
&= \mathrm{signo}(a_1b_1)\rho(X, Y).
\end{aligned}$$

En particular, se obtiene la primera igualdad en (2.58). La segunda igualdad se obtiene con el Ejercicio 3.10.
9. y 10. Estas propiedades se demuestran en el Corolario 2.6.5, más adelante. □

Si X es una variable aleatoria con $\mathrm{V}\,X > 0$ y

$$Y = a + bX, \quad \text{para ciertos} \quad a \in \mathbf{R}, \quad b \neq 0, \tag{2.59}$$

entonces

$$\rho(X, Y) = \mathrm{signo}\, b = \pm 1.$$

El recíproco también es cierto. En particular

$$\rho(X, X) = 1 \quad \text{y} \quad \rho(X, -X) = -1.$$

En contraste, si X y Y son variables aleatorias independientes (no necesariamente discretas), entonces

$$\rho(X, Y) = \mathrm{C}(X, Y) = 0.$$

De hecho

$$\mathrm{C}(X, Y) = \mathrm{E}\,XY - \mathrm{E}\,X \cdot \mathrm{E}\,Y = \mathrm{E}\,X \cdot \mathrm{E}\,Y - \mathrm{E}\,X \cdot \mathrm{E}\,Y = 0.$$

Cabe recordar que el recíproco no es cierto; como se describe en el Ejemplo 2.4.5.

Sea X una variable aleatoria de distribución binomial de parámetros $n \leq 1$ y $0 \leq p \leq 1$. Considere el número de fracasos $Y = n - X$. Entonces, Y es una variable aleatoria de distribución binomial de parámetros n y $1-p$. Las variables aleatorias X y Y están negativamente correlacionadas. De hecho, el coeficiente correlación es igual a -1. Se obtendrá la covarianza de X y Y.

$$\begin{aligned}
\mathrm{C}(X, Y) &= \mathrm{E}[XY] - \mathrm{E}\,X\,\mathrm{E}\,Y = \mathrm{E}[X(n - X)] - np \cdot n(1 - p). \\
&= \mathrm{E}[X(n - X)] - n^2p(1 - p).
\end{aligned}$$

Al aplicar la linealidad de la esperanza, se tiene

$$\begin{aligned}
\mathrm{E}\,X(n - X) &= \mathrm{E}[nX - X^2] = n\,\mathrm{E}\,X - \mathrm{E}\,X^2 \\
&= n \cdot np - [\mathrm{V}\,X + (\mathrm{E}\,X)^2] = n^2p - [np(1 - p) + (np)^2] \\
&= np[n - (1 - p) - np] = np[n - 1 - p(n - 1)] \\
&= np(n - 1)(1 - p) = n(n - 1)p(1 - p).
\end{aligned}$$

Por lo que

$$\begin{aligned} \mathrm{C}(X,Y) &= n(n-1)p(1-p) - n^2p(1-p) \\ &= np(1-p)(n-1-n) \\ &= -np(1-p) < 0. \end{aligned}$$

Se invita al lector deducir el coeficiente de correlación; véase Ejercicio 2.9.

2.6. Momentos y funciones generadoras

En esta sección se definirán los momentos centrales y no centrales de una variable aleatoria. Se definirán también las funciones generadoras de momentos y de probabilidades. Se obtendrán las propiedades básicas de dichos conceptos.

Sea X una variable aleatoria discreta, con función de densidad $f(x)$. Si

$$\mathrm{E}\,|X|^r < \infty,$$

con $r > 0$, entonces el r-ésimo *momento no central* de X se define como

$$\mathrm{E}\,X^r = \sum_x x^r f(x).$$

Así mismo, el r-ésimo *momento central* es de la variable aleatoria X

$$\mathrm{E}(X-\mu)^r = \sum_x (x-\mu)^r f(x).$$

los momentos de una variable aleatoria no determinan su distribución, incluso pueden no existir. Sin embargo, estos tienen información cualitativa de la distribución, especialmente los primeros cuatro, ya sean centrales o no. Por ejemplo, la media poblacional μ, que es el primer momento no central, representa al centro de la población. El segundo momento central corresponde a la varianza poblacional, el cual que es una medida de dispersión de la población al rededor de la media. El tercer momento central provee información del sesgo o asimetría de la población mientras que la curtosis es modulada por el cuarto momento central.

Por el Ejercicio 2.13, un momento central está bien definido si y sólo si lo está su contraparte momento no central:

$$\mathrm{E}\,|X-\mu|^r < \infty \quad \Leftrightarrow \quad \mathrm{E}\,|X|^r < \infty; \quad r \geq 1.$$

Por otro lado, si la variable aleatoria X tiene momento no central de orden s, entonces tiene todos los momentos no centrales de orden r, con $0 < r \leq s$:

$$\mathrm{E}\,X^s < \infty, \quad \text{implica} \quad \mathrm{E}\,X^r < \infty; \quad \text{para todo} \quad 0 < r \leq s.$$

Para verificar lo anterior, note que

$$\begin{aligned} |x|^r &\leq \begin{cases} |x|^s & |x| \geq 1 \\ 1 & |x| < 1 \end{cases} \\ &\leq |x|^s + 1. \end{aligned}$$

Por lo que

$$\mathrm{E}\,|X|^r \leq \mathrm{E}(|X|^s + 1) = \mathrm{E}\,|X|^s + 1 < \infty.$$

En particular, si $\mathrm{E}\,X^2 < \infty$, entonces $\mathrm{E}\,|X| < \infty$ y está bien definida la esperanza de X.

A continuación, se demostrarán cuatro desigualdades de momentos de variables aleatorias: Rogers Hölder, Cauchy Bunyakovsky Schwarz, Minkovski y Lyapunov. Estas cuatro desigualdades son equivalentes. Aquí sólo se deducirá la suficiencia de la desigualdad de Rogers Hölder para con las demás.

Teorema 2.6.1 (desigualdad de Rogers Hölder)**.** Sean X y Y dos variables aleatorias tales que $\mathrm{E}\,|X|^{1/\alpha}, \mathrm{E}\,|Y|^{1/(1-\alpha)} < \infty$, con $0 < \alpha < 1$. Entonces

$$\mathrm{E}\,|XY| \leq (\mathrm{E}\,|X|^{1/\alpha})^{\alpha}(\mathrm{E}\,|Y|^{1/(1-\alpha)})^{1-\alpha}. \tag{2.60}$$

Demostración. Si $\mathrm{E}\,|X| = 0$, por el Teorema 2.4.4.4a, se tiene $X = 0$, con probabilidad uno. Lo que implica $|XY| = |X|^{1/\alpha} = 0$, Así que (2.60) se satisface como igualdad. Se llega al mismo resultado cuando $\mathrm{E}\,|Y| = 0$. En cambio, si $\mathrm{E}\,|X|, \mathrm{E}\,|Y| > 0$, la desigualdad (2.60) equivale a

$$\begin{aligned}\mathrm{E}\left\{\frac{|X|}{(\mathrm{E}\,|X|^{1/\alpha})^{\alpha}}\frac{|Y|}{(\mathrm{E}\,|Y|^{1/(1-\alpha)})^{1-\alpha}}\right\} &= \mathrm{E}\left\{\left(\frac{|X|^{1/\alpha}}{\mathrm{E}\,|X|^{1/\alpha}}\right)^{\alpha}\left(\frac{|Y|^{1/(1-\alpha)}}{\mathrm{E}\,|Y|^{1/(1-\alpha)}}\right)^{1-\alpha}\right\}\\ &= \mathrm{E}\{U^{\alpha}V^{1-\alpha}\} \leq 1,\end{aligned}$$

donde $U = |X|^{1/\alpha}/\,\mathrm{E}\,|X|^{1/\alpha}$ y $V = |Y|^{1/(1-\alpha)}/\,\mathrm{E}\,|Y|^{1/(1-\alpha)}$. Por la desigualdad de Young (30) del Apéndice, se tiene

$$U^{\alpha}V^{1-\alpha} \leq \alpha U + (1-\alpha)V.$$

Entonces

$$\mathrm{E}\{U^{\alpha}V^{1-\alpha}\} \leq \mathrm{E}\{\alpha U + (1-\alpha)V\} = \alpha\,\mathrm{E}\,U + (1-\alpha)\,\mathrm{E}\,V = \alpha + (1-\alpha) = 1.$$

□

Teorema 2.6.2 (desigualdad de Lyapunov)**.** Sea X una variable aleatoria. Entonces

$$(\mathrm{E}\,|X|^r)^{1/r} \leq (\mathrm{E}\,|X|^s)^{1/s}\,; \quad \text{para} \quad 0 < r \leq s. \tag{2.61}$$

Demostración. Sin pérdida de generalidad, considere que X es una variable aleatoria no negativa. Al aplicar la desigualdad de Rogers Hölder (2.60), con $Y = 1$ y $\alpha = r/s$, se tiene

$$\mathrm{E}\,X \leq (\mathrm{E}\,X^{s/r})^{r/s}; \quad \text{para} \quad 0 < r \leq s.$$

Ahora aplique esta desigualdad para X^r:

$$\begin{aligned}\mathrm{E}\,X^r &\leq \{\mathrm{E}(X^r)^{s/r}\}^{r/s} = \{\mathrm{E}\,X^s\}^{r/s},\\ (\mathrm{E}\,X^r)^{1/r} &\leq (\mathrm{E}\,X^s)^{1/s}.\end{aligned}$$

□

Teorema 2.6.3 (desigualdad de Minkovski)**.** Sean X y Y dos variables aleatorias con momento de orden $r \geq 1$. Entonces, $X+Y$ tiene también momentos de orden r. Además

$$(\mathrm{E}\,|X+Y|^r)^{1/r} \leq (\mathrm{E}\,|X|^r)^{1/r} + (\mathrm{E}\,|Y|^r)^{1/r} < \infty.$$

Demostración. El caso $r=1$ se resuelve con la desigualdad del triángulo:

$$\begin{aligned} |X+Y| &\leq |X| + |Y|, \\ \mathrm{E}\,|X+Y| &\leq \mathrm{E}\,|X| + \mathrm{E}\,|Y|. \end{aligned}$$

Para $r>1$, sin pérdida de generalidad, considere $X, Y \geq 0$; con probabilidad uno. Note que

$$\begin{aligned} X+Y &\leq 2\max(X,Y), \\ (X+Y)^r &\leq 2^r \max(X^r, Y^r) \leq 2^r(X^r + Y^r), \\ \mathrm{E}(X+Y)^r &\leq 2^r(\mathrm{E}\,X^r + \mathrm{E}\,Y^r) < \infty. \end{aligned}$$

Al aplicar la desigualdad de Rogers Hölder (2.60), con $\alpha = 1/r$ y $1-\alpha = (r-1)/r$, se tiene

$$\begin{aligned} \mathrm{E}(X+Y)^r &= \mathrm{E}(X+Y)(X+Y)^{r-1} \\ &= \mathrm{E}\,X(X+Y)^{r-1} + \mathrm{E}\,Y(X+Y)^{r-1} \\ &\leq (\mathrm{E}\,X^r)^{1/r}[\mathrm{E}(X+Y)^{(r-1)r/(r-1)}]^{(r-1)/r} + (\mathrm{E}\,Y^r)^{1/r}[\mathrm{E}(X+Y)^{(r-1)r/(r-1)}]^{(r-1)/r} \\ &= [(\mathrm{E}\,X^r)^{1/r} + (\mathrm{E}\,Y^r)^{1/r}][\mathrm{E}(X+Y)^r]^{(r-1)/r}. \end{aligned}$$

Así

$$(\mathrm{E}\,|X+Y|^r)^{1/r} = \frac{\mathrm{E}(X+Y)^r}{[\mathrm{E}(X+Y)^r]^{(r-1)/r}} \leq (\mathrm{E}\,|X|^r)^{1/r} + (\mathrm{E}\,|Y|^r)^{1/r}.$$

□

Teorema 2.6.4 (desigualdad de Cauchy Bunyakovsky Schwarz)**.** Sean X y Y dos variables aleatorias con segundo momento finito. Entonces

$$(\mathrm{E}[XY])^2 \leq \mathrm{E}\,X^2\,\mathrm{E}\,Y^2. \tag{2.62}$$

Esta expresión es una igualdad, si y sólo si se cumple cualquiera de las siguientes tres afirmaciones

$$X = 0, \quad Y = 0 \quad \text{o} \quad Y = bX,$$

para cierta constante real b.

Las versiones deterministas, para sumas finitas e integrales, de la desigualdad de Cauchy Bunyakovsky Schwarz se muestran en (27) y (28) del Apéndice.

Demostración. La desigualdad de Cauchy Bunyakovsky Schwarz es una consecuencia de la de Rogers Hölder (2.60), con $\alpha = 1/2$:

$$\mathrm{E}\,|XY| \leq (\mathrm{E}\,|X|^2)^{1/2}(\mathrm{E}\,|Y|^2)^{1/2}.$$

Sin embargo, para el caso de igualdad, se ofrece una prueba alternativa. Si $\mathrm{E}\, X^2 = 0$, por el Teorema 2.4.4.4a, esto equivale a $X^2 = X = 0$. Por cual, se cumple la igualdad en (2.62). Por el contrario, cuando $\mathrm{E}\, X^2 > 0$, defina la función auxiliar

$$\begin{aligned} g(u) &= \mathrm{E}(Y - uX)^2 \\ &= \mathrm{E}[Y^2 - 2uXY + u^2X^2] \\ &= \mathrm{E}\, Y^2 - 2u\, \mathrm{E}[XY] + u^2\, \mathrm{E}\, X^2 \geq 0; \quad \text{para} \quad u \in \mathbf{R}. \end{aligned}$$

Esta es una función cuadrática no negativa, estrictamente convexa, que alcanza su valor mínimo en

$$u^* = \frac{\mathrm{E}[XY]}{\mathrm{E}\, X^2}. \tag{2.63}$$

De hecho

$$\begin{aligned} g'(u) &= -2\, \mathrm{E}[XY] + 2u\, \mathrm{E}\, X^2 = 2\, \mathrm{E}\, X^2 \left(u - \frac{\mathrm{E}[XY]}{\mathrm{E}\, X^2}\right), \\ g''(u) &= 2\, \mathrm{E}\, X^2 > 0. \end{aligned}$$

Dicho óptimo satisface

$$\begin{aligned} 0 \leq g(u^*) &= \mathrm{E}\, Y^2 - 2u^*\, \mathrm{E}[XY] + u^{*2}\, \mathrm{E}\, X^2 \\ &= \mathrm{E}\, Y^2 - 2\left(\frac{\mathrm{E}[XY]}{\mathrm{E}\, X^2}\right) \mathrm{E}[XY] + \left(\frac{\mathrm{E}[XY]}{\mathrm{E}\, X^2}\right)^2 \mathrm{E}\, X^2 \\ &= \mathrm{E}\, Y^2 + (-2+1)\frac{(\mathrm{E}[XY])^2}{\mathrm{E}\, X^2} \\ &= \frac{\mathrm{E}\, X^2\, \mathrm{E}\, Y^2 - (\mathrm{E}[XY])^2}{\mathrm{E}\, X^2}. \end{aligned}$$

Así, se obtiene (2.62). Si además $Y = bX$, entonces

$$\begin{aligned} (\mathrm{E}[XY])^2 &= (\mathrm{E}[X(bX)])^2 = b^2(\mathrm{E}\, X^2)^2 \\ &= b^2\, \mathrm{E}\, X^2\, \mathrm{E}\, X^2 = \mathrm{E}\, X^2\, \mathrm{E}[(bX)^2] \\ &= \mathrm{E}\, X^2\, \mathrm{E}\, Y^2. \end{aligned}$$

Recíprocamente, si se cumple (2.62) como igualdad, con X no idénticamente cero, entonces

$$0 \leq \mathrm{E}(Y - u^*X)^2 = g(u^*) = \frac{\mathrm{E}\, X^2\, \mathrm{E}\, Y^2 - (\mathrm{E}[XY])^2}{\mathrm{E}\, X^2} = 0,$$

con u^* dado en (2.63). Así

$$\mathrm{E}(Y - u^*X)^2 = 0.$$

De nuevo, por el Teorema 2.4.4.4.a, se tiene

$$\begin{aligned} (Y - u^*X)^2 &= 0, \\ Y - u^*X &= 0, \\ Y &= u^*X. \end{aligned}$$

□

La desigualdad de Cauchy Bunyakovsky Schwarz (2.62) implica

$$\begin{aligned}\{\mathrm{E}[(X-\mathrm{E}\,X)(Y-\mathrm{E}\,Y)]\}^2 &\leq \mathrm{E}(X-\mathrm{E}\,X)^2\,\mathrm{E}(Y-\mathrm{E}\,Y)^2,\\ [\mathrm{C}(X,Y)]^2 &\leq \mathrm{V}\,X\,\mathrm{V}\,Y.\end{aligned}$$

Si $\mathrm{V}\,X, \mathrm{V}\,Y > 0$, se tiene además

$$\begin{aligned}-\sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y} \leq\ & \mathrm{C}(X,Y) && \leq \sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y},\\ -1\ \leq\ & \frac{\mathrm{C}(X,Y)}{\sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y}} && \leq 1,\\ -1\ \leq\ & \rho(X,Y) && \leq 1.\end{aligned}$$

El siguiente corolario formaliza estas afirmaciones.

Corolario 2.6.5. Sean X y Y dos variables aleatorias con segundos momentos finitos. Entonces

$$|\,\mathrm{C}(X,Y)| \leq \sqrt{\mathrm{V}\,X\,\mathrm{V}\,Y}. \tag{2.64}$$

Si además $\mathrm{V}\,X, \mathrm{V}\,Y > 0$, entonces

$$-1 \leq \rho(X,Y) \leq 1.$$

La expresión (2.64) es igualdad si y sólo si, X o Y son constantes o

$$Y = a + bX,$$

para ciertas constantes $a \in \mathbf{R}$ y $b \neq 0$. Si además $\mathrm{V}\,X > 0$, entonces

$$\rho(X,Y) = \operatorname{signo} b = \pm 1.$$

La *función generadora de momentos* (*FGM)* de una variable aleatoria X es

$$M(t) = \mathrm{E}\,e^{tX}; \quad t \in \mathbf{R}. \tag{2.65}$$

Su dominio es el conjunto de números reales donde sea finita la esperanza: $\mathrm{E}\,e^{tX} < \infty$. Cuando menos, esta función está definida en $t = 0$:

$$M(0) = \mathrm{E}\,e^{0X} = \mathrm{E}\,1 = 1.$$

Siempre que exista, la función generadora de momentos $M(t)$ es única. Recíprocamente, esta función caracteriza a la distribución de una variable X, en cualquiera de los siguientes dos casos:

(a) $M(t)$ está bien definida en un intervalo abierto que contiene a cero, como $|t| < h$; para cierto $h > 0$. En este caso, como lo afirma el siguiente teorema, $M(t)$ es una función suave en la vecindad $(-h, h)$, en el sentido de que es infinitamente diferenciable. Además, son finitos todos los momentos de la variable aleatoria X.

(b) La variable aleatoria X es no negativa; así que $M(t)$ está bien definida cuando menos para $t \leq 0$. Si bien esta función es suave en $(-\infty, 0)$, podría no estar definida para $t > 0$ o que la variable aleatoria X no tenga sus momentos. Por ejemplo, la distribución log normal posee todos su momentos, mas sin embargo $\mathrm{E}\, e^{tX} = \infty$; para $t > 0$. Otro ejemplo es el del Ejercicio 2.11.12, que describe a una variable aleatoria no negativa sin momentos con función generadora de momentos definida sólo para $t \leq 0$.

Salvo que se mencione lo contrario, en este libro se asume el primer caso.

Teorema 2.6.6. Sea X una variable aleatoria con función generadora de momentos $M(t) = \mathrm{E}\, e^{tX}$, definida para $t \in (-t_0, t_0)$; para cierto $t_0 > 0$. Entonces

1. La variable aleatoria X tiene todos sus momentos:

$$E|X|^n < \infty \quad \text{para} \quad n \geq 0. \tag{2.66}$$

2. $M(t)$ es infinitamente diferenciable, con

$$M^{(n)}(t) = \mathrm{E}[X^n e^{tX}]; \quad n = 0, 1, 2, \ldots \tag{2.67}$$

3.
$$\mathrm{E}\, X^n = M^{(n)}(0); \quad \text{para} \quad n = 0, 1, \ldots \tag{2.68}$$

Demostración. 1. Por simplicidad, asuma que X es una variable aleatoria no negativa: $P(X \geq 0) = 1$. De hecho, note que

$$M_{|X|}(t) = \mathrm{E}\, e^{t|X|} \leq \mathrm{E}(e^{tX} + e^{-tX}) = M(t) + M(-t) < \infty; \quad t \in (-t_0, t_0).$$

Así
$$\mathrm{E}\left[\frac{t^n}{n!} X^n\right] \leq \mathrm{E}\left[1 + tX + \cdots + \frac{t^n}{n!} X^n + \cdots\right] = \mathrm{E}\, e^{tX} < \infty. \tag{2.69}$$

Por lo que $\mathrm{E}\, X^n < \infty$; para $t \in (-t_0, t_0)$ y $n = 0, 1, \ldots$.
2. Se verificará (2.67) por inducción matemática. Para $n = 1$, resulta

$$\frac{M(t+h) - M(t)}{h} = \frac{\mathrm{E}\, e^{(t+h)X} - \mathrm{E}\, e^{tX}}{h} = \mathrm{E}\, e^{tX} \frac{e^{hX} - 1}{h}.$$

Por la expansión en serie de Taylor [13, Teorema 19.4], se tiene

$$e^{hX} = 1 + e^{hY}(hX),$$

para cierta variable aleatoria $0 \leq Y \leq X$. Así

$$\begin{aligned}
\frac{M(t+h) - M(t)}{h} &= \mathrm{E}\, e^{tX}(X e^{hY}) = \mathrm{E}\, X e^{tX + hY} \\
&\leq \mathrm{E}\, X e^{(t+|h|)X} \leq (\mathrm{E}\, X^2)^{1/2} (\mathrm{E}\, e^{2(t+|h|)X})^{1/2} \\
&< \infty; \quad \text{para} \quad 2|t + |h|| < t_0.
\end{aligned}$$

Las desigualdades penúltima y última se deben a la de Cauchy Bunyakovsky Schwarz (2.62) y a (2.66). Por el Teorema de convergencia dominada (31) del Apéndice, se tiene

$$M'(t) = \lim_{h\to 0} \frac{M(t+h) - M(t)}{h} = \lim_{h\to 0} \mathrm{E}\, X e^{tX+hY} = \mathrm{E} \lim_{h\to 0} X e^{tX+hY} = \mathrm{E}\, X e^{tX};$$

para $t \in (-t_0, t_0)$. De manera similar, al asumir la hipótesis de inducción, resulta

$$\begin{aligned}\frac{M^{(n)}(t+h) - M^{(n)}(t)}{h} &= \frac{\mathrm{E}\, X^n e^{(t+h)X} - \mathrm{E}\, X^n e^{tX}}{h} = \mathrm{E}\, X^n e^{tX} \frac{e^{hX} - 1}{h} \\ &= \mathrm{E}\, X^n e^{tX} (X e^{hY}) = \mathrm{E}\, X^{n+1} e^{tX+hY} \\ &\leq \mathrm{E}\, X^{n+1} e^{(t+|h|)X} < \infty\end{aligned}$$

y

$$\begin{aligned}M^{(n+1)}(t) &= \lim_{h\to 0} \frac{M^{(n)}(t+h) - M^{(n)}(t)}{h} = \lim_{h\to 0} \mathrm{E}\, X^{n+1} e^{tX+hY} \\ &= \mathrm{E} \lim_{h\to 0} X^{n+1} e^{tX+hY} = \mathrm{E}\, X^{n+1} e^{tX}; \quad t \in (-t_0, t_0).\end{aligned}$$

3. La relación (2.68) se obtiene al evaluar (2.67) en $t = 0$. $\square$

Sea X una variable aleatoria discreta, con función de densidad $f(x) = P(X = x)$; para $x = 0, 1, 2, \ldots$, y función generadora de momentos

$$M(t) = \mathrm{E}\, e^{tX} = \sum_{x=0}^{\infty} e^{tx} f(x); \quad t \in \mathbf{R}.$$

Las derivadas de esta función se determinan por (2.67), en donde los operadores esperanza y derivada son intercambiables. Por ejemplo

$$\begin{aligned}M'(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{E}\, e^{tX} = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{x=0}^{\infty} e^{tx} f(x) \\ &= \sum_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} \{e^{tx} f(x)\} = \sum_{x=0}^{\infty} x e^{tx} f(x) \\ &= \mathrm{E}[X e^{tX}].\end{aligned}$$

Análogamente, la segunda derivada resulta

$$\begin{aligned}M''(t) &= \frac{\mathrm{d}^2}{\mathrm{d}t^2} \mathrm{E}\, e^{tX} = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{E}[X e^{tX}] \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{x=0}^{\infty} x e^{tx} f(x) = \sum_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} \{x e^{tx} f(x)\} \\ &= \sum_{x=0}^{\infty} x^2 e^{tx} f(x) = \mathrm{E}[X^2 e^{tX}].\end{aligned}$$

El mismo procedimiento aplica a la tercera derivada

$$\begin{aligned} M^{(3)}(t) &= \frac{\mathrm{d}^3}{\mathrm{d}t^3}\,\mathrm{E}[e^{tX}] = \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}\,X^2 e^{tX} \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_{x=0}^{\infty} x^2 e^{tx} f(x) = \sum_{x=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t}\{x^2 e^{tx} f(x)\} \\ &= \sum_{x=0}^{\infty} x^3 e^{tx} f(x) = \mathrm{E}[X^3 e^{tX}]. \end{aligned}$$

Así, se ratifica (2.67):

$$M^{(k)}(t) = \frac{\mathrm{d}^k}{\mathrm{d}t^k}\,\mathrm{E}\,e^{tX} = \mathrm{E}\,\frac{\mathrm{d}^k}{\mathrm{d}t^k} e^{tX} = \mathrm{E}[X^k e^{tX}], \quad \text{para} \quad k = 0, 1, \ldots$$

Al evaluar en $t = 0$, se obtienen los momentos no centrales (2.68):

$$M^{(k)}(0) = \mathrm{E}\,X^k e^{(0)X} = \mathrm{E}\,X^k, \quad k = 0, 1, \ldots$$

Los siguientes dos teoremas muestran las propiedades básicas de la función generadora de momentos. Aquí se asume que esta función está bien definida para cada variable aleatoria involucrada.

Teorema 2.6.7. Sean X y Y dos variables aleatorias con respectivas funciones generadoras de momentos $M(t) = \mathrm{E}\,e^{tX}$ y $N(t) = \mathrm{E}\,e^{tY}$; para $t \in (-t_0, t_0)$ y cierto $t_0 > 0$. Si $M(t) = N(t)$, entonces $X = Y$ en distribución.

Demostración. Se probará este resultado sólo para el caso discreto, en donde los conjuntos de valores posibles de las variables aleatorias X y Y son subconjuntos de los enteros no negativos. En este sentido

$$\begin{aligned} 0 = N(t) - M(t) &= \mathrm{E}\,e^{tY} - \mathrm{E}\,e^{tX} \\ &= \sum_{x=0}^{\infty} e^{tx} f_Y(x) - \sum_{x=0}^{\infty} e^{tx} f_X(x) = \sum_{x=0}^{\infty} e^{tx}[f_Y(x) - f_X(x)] \\ &= \sum_{x=0}^{\infty} a_x b^x, \end{aligned}$$

donde $a_x = f_Y(x) - f_X(x)$ y $b = e^t > 0$. La última expresión es un polinomio respecto de b de grado infinito, el cual es idénticamente cero. Por lo que todos sus coeficientes deben ser cero. Se concluye

$$f_Y(x) = f_X(x); \quad x = 0, 1, \ldots$$

□

Teorema 2.6.8. Considere las variables aleatorias $X, Y, X_1, \ldots, X_n$; $n \geq 1$, con funciones generadora de momentos definidas en cierto intervalo abierto que contiene a cero. Entonces

1. Para cualesquier constantes reales a y b

$$M_{a+bX}(t) = e^{at} M_X(bt).$$

2. Las variables aleatorias X y Y son independientes, si y sólo si

$$\mathrm{E}\, e^{sX+tY} = \mathrm{E}\, e^{sX}\, \mathrm{E}\, e^{tY} = M_X(s) M_Y(t); \quad \text{para todo} \quad s, t. \tag{2.70}$$

En tal caso

$$M_{X+Y}(t) = M_X(t) M_Y(t). \tag{2.71}$$

3. Las variables aleatorias $X_1, \ldots, X_n$ son independientes, si y sólo si

$$\mathrm{E}\, e^{t_1 X_1 + \cdots + t_n X_n} = \mathrm{E}\, e^{t_1 X_1} \cdots \mathrm{E}\, e^{t_n X_n} = M_{X_1}(t_1) \cdots M_{X_n}(t_n); \tag{2.72}$$

para todo $t_1, \ldots, t_n$. En tal caso

$$M_{X_1+\cdots+X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t). \tag{2.73}$$

Si además tales variables aleatorias son idénticamente distribuidas, entonces

$$M_{X_1+\cdots+X_n}(t) = [M_{X_1}(t)]^n.$$

Demostración. Por simplicidad, asuma que las funciones generadoras de momentos de las variables aleatorias invocadas en este teorema están bien definidas en la recta real.
1. Para $t \in \mathbf{R}$, se tiene

$$M_{a+bX}(t) = \mathrm{E}\, e^{t(a+bX)} = \mathrm{E}[e^{at} e^{btX}] = e^{at}\, \mathrm{E}\, e^{(bt)X} = e^{at} M_X(bt).$$

2. Si son independientes las variables aleatorias X y Y, entonces también lo serán las variables aleatorias e^{sX} y e^{tY}; véase Ejercicio 2.19. Por lo que

$$\mathrm{E}\, e^{sX+tY} = \mathrm{E}[e^{sX} e^{tY}] = \mathrm{E}\, e^{sX}\, \mathrm{E}\, e^{tY} = M_X(s) M_Y(t).$$

Para verificar el recíproco, considere la función generadora de momentos del vector aleatorio (X, Y), la cual está dada por el miembro derecho de (2.70):

$$M_{X,Y}(s, t) = \mathrm{E}\, e^{sX+tY}; \quad \text{para} \quad s, t \in \mathbf{R}. \tag{2.74}$$

Esta función es única y determina la distribución conjunta de (X, Y). Sea (U, V) un vector de variables aleatorias independientes, tales que $U \stackrel{d}{=} X$ y $V \stackrel{d}{=} Y$. Entonces

$$M_{U,V}(s, t) = \mathrm{E}\, e^{sU+tV} = \mathrm{E}\, e^{sU}\, \mathrm{E}\, e^{tV} = M_U(s) M_V(t) = M_X(s) M_Y(t) = M_{X,Y}(s, t).$$

En la última igualdad se usó (2.70). Por lo tanto, $(U, V) \stackrel{d}{=} (X, Y)$; que implica la independencia de X y Y. En particular, (2.71) se obtiene con $s = t$.
3. Si son independientes las variables aleatorias $X_1, \ldots, X_n$, entonces también lo serán las variables aleatorias $e^{t_1 X_1}, \ldots, e^{t_n X_n}$. Así

$$\mathrm{E}\, e^{t_1 X_1 + \cdots + t_n X_n} = \mathrm{E}[e^{t_1 X_1} \cdots e^{t_n X_n}] = \mathrm{E}\, e^{t_1 X_1} \cdots \mathrm{E}\, e^{t_n X_n} = M_{X_1}(t_1) \cdots M_{X_n}(t_n).$$

Para verificar el recíproco, considere la función generadora de momentos del vector aleatorio $(X_1,\dots,X_n)$, la cual está dada por el miembro derecho de (2.72):

$$M_{X_1,\dots,X_n}(t_1,\dots,t_n) = \mathrm{E}\, e^{t_1X_1+\cdots+t_nX_n}. \tag{2.75}$$

Esta función es única y determina la distribución conjunta de $(X_1,\dots,X_n)$. Sea $(U_1,\dots,U_n)$ un vector de variables aleatorias independientes, tales que $U_i \stackrel{d}{=} X_i$; para $i=1,\dots,n$. Entonces

$$\begin{aligned}
M_{U_1,\dots,U_n}(t_1,\dots,t_n) &= \mathrm{E}\, e^{t_1U_1+\cdots+t_nU_n} = \mathrm{E}\, e^{t_1U_1}\cdots \mathrm{E}\, e^{t_nU_n}\\
&= M_{U_1}(t_1)\cdots M_{U_n}(t_n) = M_{X_1}(t_1)\cdots M_{X_n}(t_n)\\
&= M_{X_1,\dots,X_n}(t_1,\dots,t_n).
\end{aligned}$$

En la última igualdad se usó (2.72). Por lo tanto, $(X_1,\dots,X_n) \stackrel{d}{=} (U_1,\dots,U_n)$; que implica la independencia de $X_1,\dots,X_n$. En particular, (2.73) se obtiene con $t_1=\cdots=t_n=t$. Si además, las n variables aleatorias tienen la misma distribución, entonces

$$M_{X_1+\cdots+X_n}(t) = M_{X_1}(t)\cdots M_{X_n}(t) = [M_{X_1}(t)]^n.$$

□

Por el Teorema 2.6.7, la distribución de la variable aleatoria X se caracteriza por la función generadora de momentos $M(t)$; siempre que esta última exista. Como consecuencia, con $M(t)$ se obtienen todos sus momentos. Por otro lado, si bien no determinan la distribución de X, los momentos aportan muchas de sus cualidades, en especial los primeros cuatro. Asuma que cuando menos, la variable aleatoria X tiene el cuarto momento finito: $\mathrm{E}\,X^4<\infty$. Este es el caso cuando esté definida su función generadora de momentos, donde aplica (2.68).

- El primer momento $\mu=\mathrm{E}\,X$, representa el centro de la población.
- El segundo momento central indica su dispersión:

  $$\sigma^2 = \mathrm{V}\,X = \mathrm{E}(X-\mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2.$$

  Recuerde que la desviación estándar es también una medida de dispersión, en la unidad de medida original:

  $$\sigma = \mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\,X}.$$

- El tercer momento central modula la simetría o asimetría de la población respecto de su media poblacional μ:

  $$\begin{aligned}
  \mathrm{E}(X-\mu)^3 &= \mathrm{E}[X^3-3\mu X^2+3\mu^2X-\mu^3]\\
  &= \mathrm{E}\,X^3 - 3\mu\,\mathrm{E}\,X^2 + 3\mu^2\,\mathrm{E}\,X - \mu^3\\
  &= \mathrm{E}\,X^3 - 3\mu\,\mathrm{E}\,X^2 + 3\mu^2\cdot\mu - \mu^3\\
  &= \mathrm{E}\,X^3 - 3\mu\,\mathrm{E}\,X^2 + 2\mu^3.
  \end{aligned}$$

  De esta manera, el *coeficiente de asimetría* o *sesgo* (*skewness*) se define como

  $$\mathrm{sesgo}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{1}{\sigma^3}\,\mathrm{E}(X-\mu)^3.$$

- El cuarto momento central establece la pesadez de sus colas y, en ocasiones, sobre la forma picuda del centro, según el *coeficiente de curtosis* (*kurtosis*)

$$\operatorname{curtosis}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^4 = \frac{1}{\sigma^4}\,\mathrm{E}(X-\mu)^4,$$

donde

$$\begin{aligned}\mathrm{E}(X-\mu)^4 &= \mathrm{E}[X^4 - 4\mu X^3 + 6\mu^2 X^2 - 4\mu^3 X + \mu^4]\\ &= \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 4\mu^3\,\mathrm{E}\,X + \mu^4\\ &= \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 4\mu^3\cdot\mu + \mu^4\\ &= \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 3\mu^4.\end{aligned}$$

En resumen, se tiene

$$\mathrm{E}(X-\mu)^2 = \mathrm{E}\,X^2 - \mu^2, \tag{2.76}$$

$$\mathrm{E}(X-\mu)^3 = \mathrm{E}\,X^3 - 3\mu\,\mathrm{E}\,X^3 + 2\mu^3, \tag{2.77}$$

$$\mathrm{E}(X-\mu)^4 = \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 3\mu^4 \tag{2.78}$$

Note que, tanto el sesgo como la curtosis de una variable aleatoria se definen en términos de valores esperados de su versión estandarizada. Esto significa que dichos indicadores son invariantes ante una transformación de localización y escala. Lo cual se analizará a continuación.

Una transformación de *localización* de la variable aleatoria X es

$$Y = a + X, \quad \text{con} \quad a \in \mathbf{R}.$$

Aquí, cada individuo de una población se trasladan a unidades, ya sea a la izquierda o a la derecha, según el signo de a. Así mismo, una transformación de *escala* es

$$Y = bX, \quad \text{con} \quad b > 0.$$

La constante b cambia la escala de la variable aleatoria. Por ejemplo, $b = 100$ cambia la medida de longitud metro a centímetro. Una transformación de *localización y escala* es

$$Y = a + bX, \quad \text{con} \quad a \in \mathbf{R} \quad \text{y} \quad b > 0. \tag{2.79}$$

En particular, si X es una variable aleatoria no constante y con varianza finita, su *estandarización* es la transformación con $a = \mu = \mathrm{E}\,X$ y $b = 1/\sigma = 1/\operatorname{d.e.}(X)$:

$$Z = \frac{X-\mu}{\sigma}.$$

La estandarización es la transformación tal que Z tenga media cero y varianza uno:

$$\mathrm{E}\,Z = \mathrm{E}\,\frac{X-\mu}{\sigma} = \frac{1}{\sigma}\,\mathrm{E}[X-\mu] = \frac{\mathrm{E}\,X-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

y

$$\mathrm{V}\,Z = \mathrm{V}\,\frac{X-\mu}{\sigma} = \frac{1}{\sigma^2}\,\mathrm{V}[X-\mu] = \frac{1}{\sigma^2}\,\mathrm{V}\,X = \frac{1}{\sigma^2}\sigma^2 = 1.$$

La estandarización es invariante ante una transformación de localización y escala. De hecho, por (2.79) se tiene

$$\begin{aligned}\mu_Y &= \mathrm{E}\,Y = \mathrm{E}[a+bX] = a + b\,\mathrm{E}\,X = a + b\mu,\\ \sigma_Y^2 &= \mathrm{V}\,Y = \mathrm{V}[a+bX] = b^2\,\mathrm{V}\,X = b^2\sigma^2\end{aligned}$$

y

$$\frac{Y-\mu_Y}{\sigma_Y} = \frac{(a+bX)-(a+b\mu)}{b\sigma} = \frac{b(X-\mu)}{b\sigma} = \frac{X-\mu}{\sigma}.$$

En consecuencia, son también invariantes el sesgo y la curtosis:

$$\mathrm{sesgo}(Y) = \mathrm{E}\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^3 = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \mathrm{sesgo}(X)$$

y

$$\mathrm{curtosis}(Y) = \mathrm{E}\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^4 = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^4 = \mathrm{curtosis}(X).$$

En cambio, cabe advertir que no se considera transformación de escala si $b < 0$ en (2.79), pues cambia el sentido de la unidad de medida. En este caso, cambia el signo del sesgo aunque se mantiene la curtosis. En resumen, para $a, b \in \mathbf{R}$, resulta

$$\begin{aligned}\mathrm{E}(a+bX) &= a + b\,\mathrm{E}\,X,\\ \mathrm{V}(a+bX) &= b^2\,\mathrm{V}\,X,\\ \mathrm{d.\,e.}(a+bX) &= |b|\,\mathrm{d.\,e.}(X),\\ \mathrm{sesgo}(a+bX) &= \mathrm{signo}\,b\cdot\mathrm{sesgo}(X),\\ \mathrm{curtosis}(a+bX) &= \mathrm{curtosis}(X)\quad (b\neq 0).\end{aligned}$$

Ejemplo 2.6.9. Considere las variables aleatorias de distribución uniforme discreta $X \sim \mathrm{U}(-1,0,1)$ y $Y \sim \mathrm{U}(-1,0,4)$. Entonces

$$\begin{aligned}\mu = \mathrm{E}\,X &= \frac{-1+0+1}{3} = 0,\\ \sigma^2 = \mathrm{V}\,X = \mathrm{E}(X-\mu)^2 = \mathrm{E}\,X^2 &= \frac{(-1)^2+0^2+1^2}{3} = \frac{2}{3} = 0.6667,\\ \sigma = \mathrm{d.\,e.}(X) = \sqrt{\mathrm{V}\,X} &= \sqrt{2/3} = 0.8165,\\ \mathrm{sesgo}(X) = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3} &= \frac{(-1)^3+0^3+1^3}{3\sigma^3} = 0,\\ \mathrm{curtosis}(X) = \frac{\mathrm{E}(X-\mu)^4}{\sigma^4} &= \frac{(-1)^4+0^4+1^4}{3(2/3)^{4/2}} = \frac{3}{2} = 1.5 \in (0,3)\end{aligned}$$

y

$$
\begin{aligned}
\mathrm{E}\,Y &= \frac{-1+0+4}{3} = 1,\\
\mathrm{V}\,Y = \mathrm{E}(Y-\mathrm{E}\,Y)^2 &= \frac{(-1-1)^2+(0-1)^2+(4-1)^2}{3} = \frac{14}{3} = 4.6667,\\
\mathrm{d.\,e.}(Y) &= \sqrt{\mathrm{V}\,Y} = \sqrt{14/3} = 2.1602,\\
\mathrm{sesgo}(Y) &= \frac{\mathrm{E}(Y-\mathrm{E}\,Y)^3}{[\mathrm{d.\,e.}(Y)]^3} = \frac{(-1-1)^3+(0-1)^3+(4-1)^3}{3(14/3)^{3/2}} = 0.5952 > 0,\\
\mathrm{curtosis}(Y) &= \frac{\mathrm{E}(Y-\mathrm{E}\,Y)^4}{[\mathrm{d.\,e.}(Y)]^4} = \frac{(-1-1)^4+(0-1)^4+(4-1)^4}{3(14/3)^{4/2}} = 1.5.
\end{aligned}
$$

La segunda distribución tiene un mayor centro, dispersión y sesgo. Ambas poblaciones coinciden en la mediana (cero) y curtosis. De hecho, por el Ejercicio 2.18, la curtosis es invariante para cualquier población de distribución uniforme discreta de tres individuos diferentes.

La *función generadora de probabilidad* (*FGP*) de una variable aleatoria X se define como

$$\phi(t) = \mathrm{E}\,t^X; \quad \text{para} \quad t > 0.$$

Esta función está bien definida siempre que $\mathrm{E}\,t^X < \infty$. Además, su dominio se puede extender para $t \leq 0$, siempre que t^X sea una variable aleatoria con esperanza finita. Por ejemplo, para el caso de una variable aleatoria no negativa, se tiene $-\infty < t^X < \infty$; para $t \in \mathbf{R}$. La relación con la función generadora de momentos es

$$\phi(t) = \mathrm{E}\,t^X = \mathrm{E}\,e^{\log t \cdot X} = M(\log t); \quad t > 0 \tag{2.80}$$

y

$$M(t) = \mathrm{E}\,e^{tX} = \mathrm{E}[(e^t)^X] = \phi(e^t); \quad t \in \mathbf{R}.$$

Por lo cual, la función generadora de probabilidad determina la distribución de la variable aleatoria X, siempre que esté definida en un intervalo abierto que contenga a uno. De hecho, con $h > 0$ fijo, $M(t)$ se define para $t \in (-h, h)$ si y sólo si, $\phi(t)$ se define para $t \in (e^{-h}, e^h)$.

Con la función generadora de probabilidad se obtienen los *momentos factoriales* de la variable aleatoria X:

$$\mathrm{E}\,X, \quad \mathrm{E}\,X(X-1), \quad \mathrm{E}\,X(X-1)(X-2), \ldots,$$

según la fórmula

$$\phi^{(k)}(1) = \mathrm{E}\,X(X-1)\cdots(X-k+1); \quad k \geq 1.$$

De hecho

$$
\begin{aligned}
\phi'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}\,t^X = \mathrm{E}\,\frac{\mathrm{d}}{\mathrm{d}t}t^X = \mathrm{E}\,Xt^{X-1},\\
\phi''(t) &= \frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\mathrm{E}\,t^X = \mathrm{E}\,X(X-1)t^{X-2},\\
&\;\vdots\\
\phi^{(k)}(t) &= \frac{\mathrm{d}^k}{\mathrm{d}t^k}\,\mathrm{E}\,t^X = \mathrm{E}\,\frac{\mathrm{d}^k}{\mathrm{d}t^k}t^X = \mathrm{E}\,X(X-1)\cdots(X-k+1)t^{X-k},\\
&\;\vdots
\end{aligned}
$$

y

$$\begin{aligned}
\phi'(1) &= \mathrm{E}\,X, \\
\phi''(1) &= \mathrm{E}\,X(X-1), \\
&\vdots \\
\phi^{(k)}(1) &= \mathrm{E}\,X(X-1)\cdots(X-k+1), \\
&\vdots
\end{aligned}$$

En particular, con las primeras cuatro derivadas de $\phi(t)$ en $t=1$, se obtienen los indicadores poblacionales de variable aleatoria X: media, varianza, sesgo y curtosis.

$$\begin{aligned}
\phi'(1) &= \mathrm{E}\,X, \\
\phi''(1) &= \mathrm{E}\,X(X-1) = \mathrm{E}[X^2 - X] = \mathrm{E}\,X^2 - \mathrm{E}\,X, \\
\phi^{(3)}(1) &= \mathrm{E}\,X(X-1)(X-2) = \mathrm{E}[X^3 - 3X^2 + 2X] = \mathrm{E}\,X^3 - 3\,\mathrm{E}\,X^2 + 2\,\mathrm{E}\,X, \\
\phi^{(4)}(1) &= \mathrm{E}\,X(X-1)(X-2)(X-3) = \mathrm{E}[X^4 - 6X^3 + 11X^2 - 6X] \\
&= \mathrm{E}\,X^4 - 6\,\mathrm{E}\,X^3 + 11\,\mathrm{E}\,X^2 - 6\,\mathrm{E}\,X.
\end{aligned}$$

Así

$$\begin{aligned}
\mathrm{E}\,X &= \phi'(1), \\
\mathrm{E}\,X^2 &= \phi''(1) + \mathrm{E}\,X = \phi''(1) + \phi'(1), \\
\mathrm{E}\,X^3 &= \phi^{(3)}(1) + 3\,\mathrm{E}\,X^2 - 2\,\mathrm{E}\,X = \phi^{(3)}(1) + 3[\phi''(1) + \phi'(1)] - 2\phi'(1) \\
&= \phi^{(3)}(1) + 3\phi''(1) + \phi'(1)
\end{aligned}$$

y

$$\begin{aligned}
\mathrm{E}\,X^4 &= \phi^{(4)}(1) + 6\,\mathrm{E}\,X^3 - 11\,\mathrm{E}\,X^2 + 6\,\mathrm{E}\,X \\
&= \phi^{(4)}(1) + 6[\phi^{(3)}(1) + 3\phi''(1) + \phi'(1)] - 11[\phi''(1) + \phi'(1)] + 6\phi'(1) \\
&= \phi^{(4)}(1) + 6\phi^{(3)}(1) + 7\phi''(1) + \phi'(1).
\end{aligned}$$

En resumen, se tiene

$$\mathrm{E}\,X = \mu = \phi'(1), \tag{2.81}$$
$$\mathrm{E}\,X^2 = \mu + \phi''(1), \tag{2.82}$$
$$\mathrm{E}\,X^3 = \mu + 3\phi''(1) + \phi^{(3)}(1), \tag{2.83}$$
$$\mathrm{E}\,X^4 = \mu + 7\phi''(1) + 6\phi^{(3)}(1) + \phi^{(4)}(1). \tag{2.84}$$

El nombre de función generadora de probabilidad se debe a que, en el caso de una variable aleatoria discreta entera no negativa, obtiene las probabilidades $f(x) = P(X = x)$. De esta manera, la función $\phi(t)$ determina la función de densidad discreta $f(x)$, y viceversa.

Sea X una variable aleatoria discreta con función de densidad $f(x) = P(X = x)$; para $x = 0, 1, \ldots$ Considere que su función generadora de probabilidad está definida en un intervalo que contiene a cero. Por simplicidad, asuma que dicho intervalo es la recta real:

$$\phi(t) = \mathrm{E}\, t^X = \sum_{x=0}^{\infty} t^x f(x); \quad \text{para} \quad t \in \mathbf{R}.$$

La correspondiente expansión en serie de Taylor es

$$\phi(t) = \sum_{x=0}^{\infty} \frac{\phi^{(x)}(0)}{x!} t^x; \quad \text{para} \quad t \in \mathbf{R}.$$

Al comparar los coeficientes de ambas sumas, se tiene

$$f(x) = \frac{\phi^{(x)}(0)}{x!}; \quad x = 0, 1, \ldots$$

Por ejemplo, para $t = 0$, se tiene

$$0^X = \begin{cases} 1 & X = 0 \\ 0 & X > 0. \end{cases}$$

Así que

$$\frac{\phi(0)}{0!} = \phi(0) = \mathrm{E}\, 0^X = 1P(X=0) + 0P(X>0) = P(X=0) = f(0).$$

Análogo al Teorema 2.6.8 para la FGM, la función generadora de probabilidad satisface las siguientes propiedades básicas.

Teorema 2.6.10. Considere las variables aleatorias $X, Y, X_1, \ldots, X_n$; $n \geq 1$, con funciones generadora de probabilidad finitas en un intervalo abierto que contiene a uno. Entonces

1. Para cualesquier constantes reales a y b, se tiene

$$\phi_{a+bX}(t) = t^a \phi_X(t^b).$$

2. Si las variables aleatorias X y Y son independientes, entonces

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

3. Si son independientes las variables aleatorias $X_1, \ldots, X_n$, entonces

$$\phi_{X_1+\cdots+X_n}(t) = \phi_{X_1}(t) \cdots \phi_{X_n}(t).$$

Si además, tales variables aleatorias son idénticamente distribuidas, entonces

$$\phi_{X_1+\cdots+X_n}(t) = [\phi_{X_1}(t)]^n.$$

A continuación, se describirán algunos ejemplos de distribuciones de probabilidad.

Distribución uniforme discreta. Sea X es una variable aleatoria con distribución uniforme discreta, $X \sim \mathrm{U}(1, \ldots, N)$; con $N \geq 1$. Al considerar su función de densidad (2.31), la función generadora de momentos es

$$M(t) = \frac{1}{N}(e^t + e^{2t} + \cdots + e^{Nt}); \quad t \in \mathbf{R}. \tag{2.85}$$

Su deducción es la siguiente

$$M(t) = \mathrm{E}\, e^{tX} = \sum_{x=0}^{\infty} e^{tx} f(x) = \sum_{x=1}^{N} e^{tx} \frac{1}{N} = \frac{1}{N} \sum_{x=1}^{N} e^{tx}; \quad t \in \mathbf{R}.$$

De la misma manera, se obtiene la función generadora de probabilidad

$$\phi(t) = \mathrm{E}\, t^X = \sum_{x=0}^{\infty} t^x f(x) = \sum_{x=1}^{N} t^x \frac{1}{N} = \frac{1}{N}(t + t^2 + \cdots + t^N); \quad t \in \mathbf{R}.$$

Con un dominio reducido, esta fórmula se obtiene al aplicar (2.80):

$$\phi(t) = \mathrm{E}\, t^X = M(\log t) = \frac{1}{N}(t + t^2 + \cdots + t^N); \quad t > 0.$$

Las primeras dos derivadas de la función $M(t)$ son

$$M'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{N}(e^t + e^{2t} + \cdots + e^{Nt})\right] = \frac{1}{N}(e^t + 2e^{2t} + \cdots + Ne^{Nt})$$

y

$$M''(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{N}(e^t + 2e^{2t} + \cdots + Ne^{Nt})\right] = \frac{1}{N}(e^t + 4e^{2t} + \cdots + N^2 e^{Nt}).$$

Así que

$$\begin{aligned}
\mathrm{E}\, X = M'(0) &= \frac{1}{N}(e^0 + 2e^{2(0)} + \cdots + Ne^{N(0)}) \\
&= \frac{1}{N}(1 + 2 + \cdots + N) = \frac{1}{N}\frac{N(N+1)}{2} \\
&= \frac{N+1}{2},
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\, X^2 = M''(0) &= \frac{1}{N}(e^0 + 4e^{2(0)} + \cdots + N^2 e^{N(0)}) \\
&= \frac{1}{N}(1 + 2^2 + \cdots + N^2) = \frac{1}{N}\frac{N(N+1)(2N+1)}{6} \\
&= \frac{(N+1)(2N+1)}{6}
\end{aligned}$$

y

$$
\begin{aligned}
\mathrm{V}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\
&= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{(N+1)[2(2N+1)-3(N+1)]}{12} \\
&= \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12}.
\end{aligned}
$$

Compare estos resultados con (2.32).

Distribución geométrica. Si X es una variable aleatoria de distribución geométrica, con función de densidad (2.19), su función generadora de momentos se obtiene como sigue

$$
\begin{aligned}
M(t) = \mathrm{E}\,e^{tX} &= \sum_{x=0}^{\infty} e^{tx} f(x) \\
&= \sum_{x=1}^{\infty} e^{tx} \cdot p(1-p)^{x-1} = p \sum_{x=1}^{\infty} e^{t[(x-1)+1]}(1-p)^{x-1} \\
&= pe^t \sum_{x=1}^{\infty} e^{t(x-1)}(1-p)^{x-1} = pe^t \sum_{z=0}^{\infty} e^{tz}(1-p)^z \\
&= pe^t \sum_{z=0}^{\infty} [e^t(1-p)]^z = pe^t \frac{1}{1-e^t(1-p)} \\
&= \frac{pe^t}{1-(1-p)e^t}.
\end{aligned}
$$

Dicha función está bien definida siempre que

$$
0 < (1-p)e^t < 1, \qquad e^t < \frac{1}{1-p}, \qquad t < -\log(1-p).
$$

En caso contrario, para $t \geq -\log(1-p)$, se tiene $\mathrm{E}\,e^{tX} = \infty$. En conclusión

$$
M(t) = \mathrm{E}\,e^{tX} = \frac{pe^t}{1-(1-p)e^t}; \quad \text{para} \quad t < -\log(1-p).
$$

La función generadora de probabilidad se obtiene de manera similar:

$$
\begin{aligned}
\phi(t) = \mathrm{E}\,t^X &= \sum_{x=0}^{\infty} t^x f(x) \\
&= \sum_{x=1}^{\infty} t^x \cdot p(1-p)^{x-1} = p \sum_{x=1}^{\infty} t^{(x-1)+1}(1-p)^{x-1} \\
&= pt \sum_{x=1}^{\infty} [t(1-p)]^{x-1} = pt \sum_{z=0}^{\infty} [t(1-p)]^z \\
&= \frac{pt}{1-(1-p)t}; \quad -\frac{1}{1-p} < t < \frac{1}{1-p}.
\end{aligned}
$$

Con un dominio restringido, la FGP se obtienen también vía (2.80):

$$\phi(t) = \mathrm{E}\, t^X = M(\log t) = \frac{pe^{\log t}}{1-(1-p)e^{\log t}} = \frac{pt}{1-(1-p)t}; \quad 0 < t < \frac{1}{1-p}.$$

Para el cálculo de los dos primeros momentos de la distribución geométrica, se usa tanto la FGM como la FGP.

$$\begin{aligned}
M(t) &= \frac{pe^t}{1-(1-p)e^t},\\
M'(t) &= \frac{[1-(1-p)e^t]\cdot pe^t + pe^t\cdot(1-p)e^t}{(1-(1-p)e^t)^2} = \frac{pe^t}{(1-(1-p)e^t)^2} = \frac{M(t)}{1-(1-p)e^t},\\
M''(t) &= \frac{(1-(1-p)e^t)M'(t) + M(t)(1-p)e^t}{(1-(1-p)e^t)^2} = \frac{M(t)+M(t)(1-p)e^t}{(1-(1-p)e^t)^2}\\
&= \frac{M(t)[1+(1-p)e^t]}{(1-(1-p)e^t)^2}.
\end{aligned}$$

Así

$$\begin{aligned}
\mathrm{E}\, X = M'(0) &= \frac{M(0)}{1-(1-p)e^0} = \frac{1}{1-(1-p)} = \frac{1}{p},\\
\mathrm{E}\, X^2 = M''(0) &= \frac{M(0)[1+(1-p)e^0]}{[1-(1-p)e^0]^2} = \frac{1+(1-p)}{[1-(1-p)]^2} = \frac{2-p}{p^2}
\end{aligned}$$

y

$$\mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = M''(0) - [M'(0)]^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.$$

Con la ruta de la FGP, la deducción es más simple

$$\begin{aligned}
\phi(t) &= \frac{pt}{1-(1-p)t},\\
\phi'(t) &= p\frac{[1-(1-p)t]+t(1-p)}{[1-(1-p)t]^2} = \frac{p}{[1-(1-p)t]^2},\\
\phi''(t) &= \frac{2p(1-p)}{[1-(1-p)t]^3}.
\end{aligned}$$

Por lo que

$$\begin{aligned}
\phi'(1) &= \frac{p}{[1-(1-p)]^2} = \frac{1}{p} = \mathrm{E}\, X,\\
\phi''(1) &= \frac{2p(1-p)}{[1-(1-p)]^3} = \frac{2(1-p)}{p^2} = \mathrm{E}\, X^2 - \mathrm{E}\, X,\\
\mathrm{E}\, X^2 &= \mathrm{E}\, X + \frac{2(1-p)}{p^2} = \frac{1}{p} + \frac{2(1-p)}{p^2} = \frac{p+2(1-p)}{p^2} = \frac{2-p}{p^2},\\
\mathrm{V}\, X &= \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

Compare estos resultados con (2.35).

Distribución binomial. Sea X una variable aleatoria de distribución binomial de parámetros n y p. Sus funciones generadoras de momentos y de probabilidad son, respectivamente

$$M(t) = \operatorname{E} e^{tX} = [pe^t + 1 - p]^n, \tag{2.86}$$
$$\phi(t) = \operatorname{E} t^X = [pt + 1 - p]^n; \quad t \in \mathbf{R}. \tag{2.87}$$

La función generadora de momentos se deduce por lo siguiente

$$\begin{aligned}
M(t) = \operatorname{E} e^{tX} &= \sum_{x=0}^{\infty} e^{tx} f(x) \\
&= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\
&= [pe^t + 1 - p]^n.
\end{aligned}$$

En la última igualdad se usó la fórmula del binomio (20) del Apéndice. Sus primeras tres derivadas son

$$\begin{aligned}
M'(t) &= npe^t[pe^t + 1 - p]^{n-1}, \\
M''(t) &= npe^t[pe^t + 1 - p]^{n-1} + n(n-1)pe^t\{pe^t\}[pe^t + 1 - p]^{n-2} \\
&= M'(t) + n(n-1)p^2e^{2t}[pe^t + 1 - p]^{n-2}, \\
M^{(3)}(t) &= M''(t) + 2n(n-1)p^2e^{2t}[pe^t + 1 - p]^{n-2} \\
&\quad + n(n-1)(n-2)p^2e^{2t}\{pe^t\}[pe^t + 1 - p]^{n-3} \\
&= M''(t) + 2n(n-1)p^2e^{2t}[pe^t + 1 - p]^{n-2} \\
&\quad + n(n-1)(n-2)p^3e^{3t}[pe^t + 1 - p]^{n-3}.
\end{aligned}$$

Entonces

$$\begin{aligned}
\operatorname{E} X &= M'(0) = npe^0[pe^0 + 1 - p]^{n-1} = np = \mu, \\
\operatorname{E} X^2 &= M''(0) = M'(0) + n(n-1)p^2 = \mu + \mu(\mu - p), \\
\operatorname{V} X &= \operatorname{E} X^2 - (\operatorname{E} X)^2 = [\mu + \mu(\mu - p)] - \mu^2 = \mu(1-p) = np(1-p),
\end{aligned}$$

$$\begin{aligned}
\operatorname{E} X^3 &= M^{(3)}(0) \\
&= M''(0) + 2n(n-1)p^2 + n(n-1)(n-2)p^3 \\
&= [\mu + \mu(\mu - p)] + 2\mu(\mu - p) + \mu(\mu - p)(\mu - 2p) \\
&= \mu + 3\mu(\mu - p) + \mu(\mu - p)(\mu - 2p), \\
\operatorname{E}(X - \mu)^3 &= EX^3 - 3\mu \operatorname{E} X^2 + 2\mu^3 \\
&= [\mu + 3\mu(\mu - p) + \mu(\mu - p)(\mu - 2p)] - 3\mu[\mu + \mu(\mu - p)] + 2\mu^3 \\
&= \mu[(1 - 3p + 2p^2) + (3 - 3p - 3 + 3p)\mu + (1 - 3 + 2)\mu^2] \\
&= \mu(1-p)(1-2p) \\
&= np(1-p)(1-2p)
\end{aligned}$$

y

$$\text{sesgo}(X) = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3} = \frac{np(1-p)(1-2p)}{(\sigma^2)^{3/2}} = \frac{np(1-p)(1-2p)}{(np(1-p))^{3/2}} = \frac{1-2p}{\sqrt{np(1-p)}}.$$

Por último

$$\begin{aligned}
M^{(4)}(t) &= M^{(3)}(t) + 2n(n-1)p^2\{2e^{2t}\}[pe^t+1-p]^{n-2} \\
&\quad + 2n(n-1)(n-2)p^2e^{2t}\{pe^t\}[pe^t+1-p]^{n-3} \\
&\quad + n(n-1)(n-2)p^3\{3e^{3t}\}[pe^t+1-p]^{n-3} \\
&\quad + n(n-1)(n-2)(n-3)p^3e^{3t}\{pe^t\}[pe^t+1-p]^{n-4} \\
&= M^{(3)}(t) + 4n(n-1)p^2e^{2t}[pe^t+1-p]^{n-2} \\
&\quad + 5n(n-1)(n-2)p^3e^{3t}[pe^t+1-p]^{n-3} \\
&\quad + n(n-1)(n-2)(n-3)p^4e^{4t}[pe^t+1-p]^{n-4}
\end{aligned}$$

y

$$\begin{aligned}
EX^4 &= M^{(4)}(0) = M^{(3)}(0) + 4n(n-1)p^2 + 5n(n-1)(n-2)p^3 + n(n-1)(n-2)(n-3)p^4 \\
&= [\mu + 3\mu(\mu-p) + \mu(\mu-p)(\mu-2p)] + 4\mu(\mu-p) \\
&\quad + 5\mu(\mu-p)(\mu-2p) + \mu(\mu-p)(\mu-2p)(\mu-3p) \\
&= \mu[(1-3p+2p^2-4p+10p^2-6p^3) + (3-3p+4-15p+11p^2)\mu \\
&\quad + (1+5-6p)\mu^2 + \mu^3] \\
&= \mu[(1-7p+12p^2-6p^3) + (7-18p+11p^2)\mu + 6(1-p)\mu^2 + \mu^3] \\
&= \mu[(1-p)(1-6p+6p^2) + (1-p)(7-11p)\mu + 6(1-p)\mu^2 + \mu^3].
\end{aligned}$$

Por el Ejercicio 2.11.16 se demuestra que el coeficiente de curtosis es

$$\text{curtosis}(X) = \frac{\mathrm{E}(X-\mu)^4}{\sigma^4} = \frac{1-6p(1-p)}{np(1-p)} + 3. \tag{2.88}$$

El coeficiente de *curtosis excedente* es

$$\text{curtosis.e.}(X) = \text{curtosis}(X) - 3 = \frac{1-6p(1-p)}{np(1-p)}.$$

Con la función generadora de probabilidad es más fácil el cálculo de los primeros cuatro momentos de la variable aleatoria X.

$$\begin{aligned}
\phi(t) = \mathrm{E}\, t^X &= \sum_{x=0}^{\infty} t^x f(x) \\
&= \sum_{x=0}^{n} t^x \cdot \binom{n}{x} p^x(1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pt)^x (1-p)^{n-x} \\
&= [pt+1-p]^n; \quad t \in \mathbf{R}.
\end{aligned}$$

Así

$$\begin{aligned}
\phi'(t) &= np(pt+1-p)^{n-1},\\
\phi''(t) &= n(n-1)p^2(pt+1-p)^{n-2},\\
\phi^{(3)}(t) &= n(n-1)(n-2)p^3(pt+1-p)^{n-3},\\
\phi^{(4)}(t) &= n(n-1)(n-2)(n-3)p^4(pt+1-p)^{n-4}
\end{aligned}$$

y

$$\begin{aligned}
\phi'(1) &= np(p+1-p)^{n-1} = np = \mu,\\
\phi''(1) &= n(n-1)p^2 = np(np-p) = \mu(\mu-p),\\
\phi^{(3)}(1) &= n(n-1)(n-2)p^3 = \mu(\mu-p)(\mu-2p),\\
\phi^{(4)}(1) &= n(n-1)(n-2)(n-3)p^4 = \mu(\mu-p)(\mu-2p)(\mu-3p).
\end{aligned}$$

Por expresiones (2.81)-(2.84), se tiene

$$\begin{aligned}
\mathrm{E}\,X &= \mu = \phi'(1) = np,\\
\mathrm{E}\,X^2 &= \mu + \phi''(1) = \mu + \mu(\mu-p),\\
\mathrm{E}\,X^3 &= \mu + 3\phi''(1) + \phi^{(3)}(1) = \mu + 3\mu(\mu-p) + \mu(\mu-p)(\mu-2p),\\
\mathrm{E}\,X^4 &= \mu + 7\phi''(1) + 6\phi^{(3)}(1) + \phi^{(4)}(1)\\
&= \mu + 7\mu(\mu-p) + 6\mu(\mu-p)(\mu-2p) + \mu(\mu-p)(\mu-2p)(\mu-3p).
\end{aligned}$$

Por (2.76)-(2.78), resulta

$$\begin{aligned}
\mathrm{V}\,X &= \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = \mu + \mu(\mu-p) - \mu^2 = \mu(1-p) = np(1-p),\\
\mathrm{E}(X-\mu)^3 &= \mathrm{E}\,X^3 - 3\,\mathrm{E}\,X\,\mathrm{E}\,X^2 + 2(\mathrm{E}\,X)^3\\
&= [\mu + 3\mu(\mu-p) + \mu(\mu-p)(\mu-2p)] - 3\mu[\mu + \mu(\mu-p)] + 2\mu^3\\
&= \mu[1 + p^2 + \mu(3-3p-3) + \mu^2(1+2)]\\
&= \phi^{(3)}(1) - 3(\mu-1)\phi''(1) + \mu(\mu-1)(2\mu-1)\\
&= \mu(\mu-p)(\mu-2p) - 3(\mu-1)\mu(\mu-p) + \mu(\mu-1)(2\mu-1)\\
&= \mu(\mu-p)[\mu-2p-3(\mu-1)+(2\mu-1)] + \mu(p-1)(2\mu-1)\\
&= \mu(\mu-p)2(1-p) - \mu(1-p)(2\mu-1)\\
&= \mu(1-p)[2(\mu-p)-(2\mu-1)]\\
&= \mu(1-p)(1-2p) = np(1-p)(1-2p)
\end{aligned}$$

y

$$\begin{aligned}
\mathrm{E}(X-\mu)^4 &= \phi^{(4)}(1) - 2(2\mu-3)\phi^{(3)}(1)\\
&\quad + (7-12\mu+6\mu^2)\phi''(1) - \mu(\mu-1)(1-3\mu+3\mu^2)\\
&= \mu(\mu-p)(\mu-2p)(\mu-3p) - 2(2\mu-3)\cdot\mu(\mu-p)(\mu-2p)\\
&\quad + (7-12\mu+6\mu^2)\cdot\mu(\mu-p) - \mu(\mu-1)(1-3\mu+3\mu^2)\\
&= \mu[(\mu-p)(\mu-2p)(\mu-3p) - 2(2\mu-3)(\mu-p)(\mu-2p)\\
&\quad + (7-12\mu+6\mu^2)(\mu-p) - (\mu-1)(1-3\mu+3\mu^2)]\\
&= \mu(1-p)[1+3(\mu-2p)(1-p)] = np(1-p)[1+3(n-2)p(1-p)].
\end{aligned}$$

El coeficiente de sesgo o asimetría es

$$\operatorname{sesgo}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\mathrm{E}(X-\mu)^3}{(\sigma^2)^{3/2}} = \frac{np(1-p)(1-2p)}{[np(1-p)]^{3/2}} = \frac{1-2p}{[np(1-p)]^{1/2}}.$$

Se confirma que el coeficiente de curtosis es (2.88):

$$\begin{aligned}\operatorname{curtosis}(X) &= \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^4 = \frac{\mathrm{E}(X-\mu)^4}{(\sigma^2)^2} \\ &= \frac{np(1-p)[1+3(n-2)p(1-p)]}{[np(1-p)]^2} = \frac{1+3(n-2)p(1-p)}{np(1-p)}.\end{aligned}$$

La distribución binomial es simétrica si y sólo si, $p = 1/2$. Su sesgo es a la derecha o positivo [izquierda o negativo] cuando $0 < p < 1/2$ $[1/2 < p < 1]$. Para $0 < p < 1$, la población es cada vez más simétrica conforme crece el número n de ensayos. Así mismo, la curtosis se aproxima a 3. De hecho

$$\lim_{n\to\infty} \operatorname{sesgo}(X) = \lim_{n\to\infty} \frac{1-2p}{[n(1-p)]^{1/2}} = 0$$

y

$$\lim_{n\to\infty} \operatorname{curtosis}(X) = \lim_{n\to\infty} \frac{1+3(n-2)(1-p)}{n(1-p)} = \frac{3(1-p)}{1-p} = 3.$$

Previa estandarización, la función de densidad binomial tiende a la función de densidad normal estándar (campana de Gauss). Cabe aclarar que esta última función es continua.

Distribución de Poisson. Si X es una variable aleatoria con distribución Poisson de media $\mu > 0$. Entonces, sus funciones generadoras son

$$M(t) = \mathrm{E}\, e^{tX} = e^{\mu(e^t-1)}; \quad t \in \mathbf{R}$$

y

$$\phi(t) = \mathrm{E}\, t^X = M(\log t) = e^{\mu(e^{\log t}-1)} = e^{\mu(t-1)}; \quad t > 0.$$

De hecho

$$\begin{aligned}M(t) = \mathrm{E}\, e^{tX} &= \sum_{x=0}^{\infty} e^{tx} f(x) \\ &= \sum_{x=0}^{\infty} e^{tx} \cdot e^{-\mu}\frac{\mu^x}{x!} = e^{-\mu}\sum_{x=0}^{\infty} (e^t)^x \frac{\mu^x}{x!} \\ &= e^{-\mu}\sum_{x=0}^{\infty} \frac{(e^t\mu)^x}{x!} = e^{-\mu} e^{e^t\mu} \\ &= e^{\mu e^t-\mu} = e^{\mu(e^t-1)}.\end{aligned}$$

Se calcularán los primeros cuatro momentos de una población con distribución de Poisson:

$$\begin{aligned}
M'(t) &= \mu e^t e^{\mu(e^t-1)} = \mu e^t M(t),\\
M''(t) &= \mu e^t(M(t)+M'(t)) = \mu e^t(M(t)+\mu e^t M(t)) = (\mu e^t+\mu^2 e^{2t})M(t),\\
M^{(3)}(t) &= (\mu e^t + 2\mu^2 e^{2t})M(t) + (\mu e^t+\mu^2 e^{2t})\mu e^t M(t)\\
&= (\mu e^t + 3\mu^2 e^{2t} + \mu^3 e^{3t})M(t),\\
M^{(4)}(t) &= (\mu e^t + 6\mu^2 e^{2t} + 3\mu^3 e^{3t})M(t) + (\mu e^t + 3\mu^2 e^{2t} + \mu^3 e^{3t})\mu e^t M(t)\\
&= (\mu e^t + 7\mu^2 e^{2t} + 6\mu^3 e^{3t} + \mu^4 e^{4t})M(t)
\end{aligned}$$

Por lo que

$$\begin{aligned}
1 &= M(0) = e^{\mu(e^0-1)} = e^0 = 1,\\
\mathrm{E}\,X &= M'(0) = \mu e^0 M(0) = \mu,\\
\mathrm{E}\,X^2 &= M''(0) = \mu e^0 + \mu^2 e^{2(0)} = \mu+\mu^2,\\
\mathrm{E}\,X^3 &= M^{(3)}(0) = \mu + 3\mu^2 + \mu^3,\\
\mathrm{E}\,X^4 &= M^{(4)}(0) = \mu + 7\mu^2 + 6\mu^3 + \mu^4
\end{aligned}$$

y

$$\begin{aligned}
\mathrm{V}\,X &= \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = (\mu+\mu^2) - \mu^2 = \mu,\\
\mathrm{E}(X-\mu)^3 &= \mathrm{E}\,X^3 - 3\mu\,\mathrm{E}\,X^2 + 2\mu^3 = (\mu+3\mu^2+\mu^3) - 3\mu(\mu+\mu^2) + 2\mu^3 = \mu,\\
\mathrm{E}(X-\mu)^4 &= \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 3\mu^4\\
&= (\mu+7\mu^2+6\mu^3+\mu^4) - 4\mu(\mu+3\mu^2+\mu^3) + 6\mu^2(\mu+\mu^2) - 3\mu^4\\
&= \mu + (7-4)\mu^2 + (6-12+6)\mu^3 + (1-4+6-3)\mu^4 = \mu + 3\mu^2.
\end{aligned}$$

Por lo que

$$\mathrm{sesgo}(X) = \frac{\mathrm{E}(X-\mu)^3}{\sigma^3} = \frac{\mu}{(\mu^{1/2})^3} = \frac{\mu}{\mu^{3/2}} = \frac{1}{\mu^{1/2}}$$

y

$$\mathrm{curtosis}(X) = \frac{\mathrm{E}(X-\mu)^4}{\sigma^4} = \frac{\mu+3\mu^2}{(\mu^{1/2})^4} = \frac{\mu+3\mu^2}{\mu^2} = 3 + \frac{1}{\mu}.$$

La estandarización de la distribución de Poisson es asintóticamente simétrica con curtosis 3:

$$\lim_{\mu\to\infty} \mathrm{sesgo}(X) = \lim_{\mu\to\infty} \frac{1}{\mu^{1/2}} = 0$$

y

$$\lim_{\mu\to\infty} \mathrm{curtosis}(X) = \lim_{\mu\to\infty}\left(3+\frac{1}{\mu}\right) = 3.$$

Esto evidencia la aproximación normal de la distribución de Poisson. Por otro lado, con la función generadora de probabilidad se obtiene un resultado similar

$$\begin{aligned}
\phi(t) &= \mathrm{E}\, t^X = M(\log t) = e^{\mu(t-1)},\\
\phi'(t) &= \mu e^{\mu(t-1)},\\
\phi''(t) &= \mu^2 e^{\mu(t-1)},\\
\phi^{(3)}(t) &= \mu^3 e^{\mu(t-1)},\\
\phi^{(4)}(t) &= \mu^4 e^{\mu(t-1)},\\
&\vdots\\
\phi^{(k)}(t) &= \mu^k e^{\mu(t-1)},\\
&\vdots
\end{aligned}$$

Así que

$$\phi^{(k)}(1) = \mu^k e^{\mu\times 0} = \mu^k; \quad \text{para} \quad k \geq 0.$$

Por expresión (2.81), se obtienen los mismos resultados vía la función generadora de momentos:

$$\begin{aligned}
\mathrm{E}\, X &= \phi'(1) = \mu,\\
\mathrm{E}\, X^2 &= \phi'(1) + \phi''(1) = \mu + \mu^2,\\
\mathrm{E}\, X^3 &= \phi'(1) + 3\phi''(1) + \phi^{(3)}(1) = \mu + 3\mu^2 + \mu^3,\\
\mathrm{E}\, X^4 &= \phi'(1) + 7\phi''(1) + 6\phi^{(3)}(1) + \phi^{(4)}(1) = \mu + 7\mu^2 + 6\mu^3 + \mu^4.
\end{aligned}$$

Por otro lado, la FGP de la distribución de Poisson se linealiza con la transformación logaritmo:

$$\log \phi(t) = \log e^{\mu(t-1)} = \mu(t-1); \quad t > 0.$$

Este resultado permite validar empíricamente la distribución de Poisson de una población. La versión empírica de $\log \phi(t)$ tendría una tendencia lineal con pendiente cercana a $\hat{\mu}$ e intercepto $-\hat{\mu}$. Aquí $\hat{\mu}$ denota la media muestral. Si $X_1, \ldots, X_n$, es una muestra aleatoria de una población con distribución de Poisson de media μ, la *función generadora de probabilidad empírica* es

$$\hat{\phi}(t) = \frac{1}{n}\sum_{i=1}^{n} t^{X_i}; \quad t \geq 0.$$

Entonces, la gráfica de la función

$$\log \hat{\phi}(t) = \log\left(\frac{1}{n}\sum_{i=1}^{n} t^{X_i}\right); \quad t \geq 0,$$

es aproximadamente una línea recta con pendiente $\hat{\mu}$ e intercepto $-\hat{\mu}$. Cabe mencionar que, en general, la FGP empírica $\hat{\phi}(t)$ es un estimador insesgado y consistente de $\phi(t)$, puesto que

$$\mathrm{E}\, \hat{\phi}(t) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} t^{X_i}\right] = \frac{n}{n}\,\mathrm{E}\, t^X = \phi(t)$$

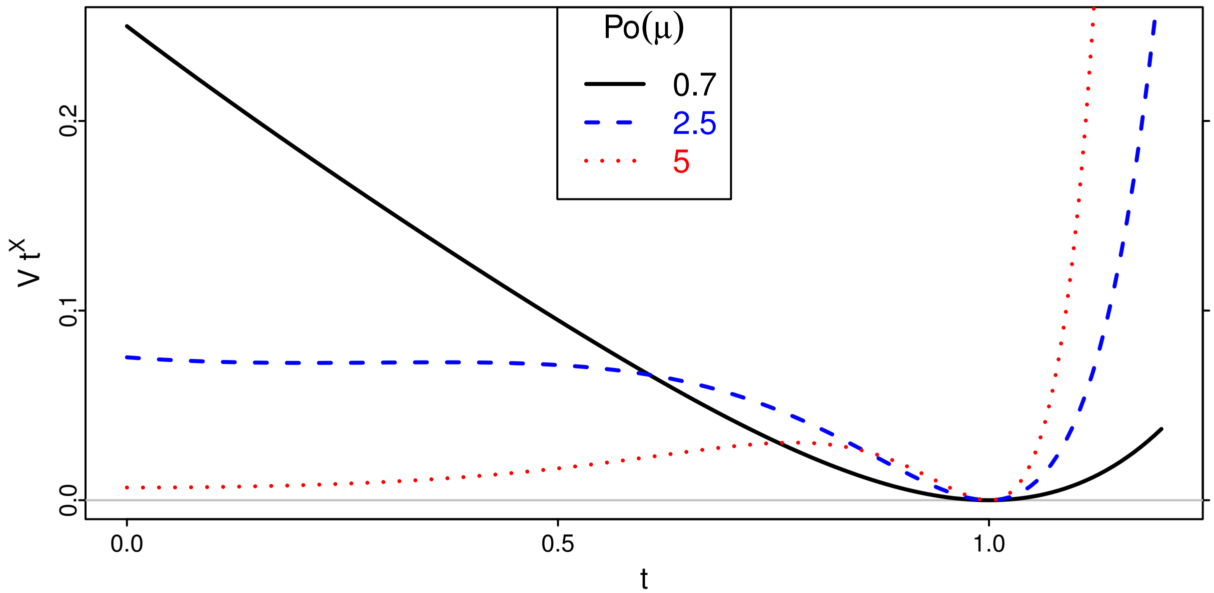


Figura 2.13. Gráficas de varianza de t^X; $t > 0$, para variables aleatorias de distribución de Poisson.

y

$$\mathrm{V}\,\hat{\phi}(t) = \mathrm{V}\left[\frac{1}{n}\sum_{i=1}^{n} t^{X_i}\right] = \frac{n}{n^2}\,\mathrm{V}\,t^X = \frac{1}{n}\,\mathrm{V}\,t^X,$$

con

$$\mathrm{V}\,t^X = \mathrm{E}(t^X)^2 - (\mathrm{E}\,t^X)^2 = \mathrm{E}\,t^{2X} - (\mathrm{E}\,t^X)^2 = \phi(t^2) - \phi^2(t) < \infty.$$

En particular, para la distribución de Poisson, se tiene

$$\mathrm{V}\,t^X = \phi(t^2) - \phi^2(t) = e^{\mu(t^2-1)} - e^{2\mu(t-1)}; \quad t \geq 0.$$

La gráfica de la varianza de t^X; para $t \in \mathbf{R}$, se muestra en la Figura 2.13, para los ejemplos de medias $\mu = 0.7, 2.5, 5$; de curvas respectivas continua negra, segmentada azul y punteado rojo. Observe que la varianza de t^X está acotada para $t \in [0, 1]$, mientras que alcanza su mínimo en $t = 1$, en donde se anula. Además, la varianza tiende rápidamente a ∞ cuando t crece, especialmente para μ grande. Por estas propiedades, en [10] sugieren la gráfica de la log función generadora de probabilidad empírica $\log\hat{\phi}(t)$, para $t \in [0, 1]$.

Por otro lado, una importante aplicación de las funciones generadoras es conocer la distribución de la suma de dos o más variables aleatorias independientes e idénticamente distribuidas. Por ejemplo, se demuestra fácilmente que la suma de dos variables aleatorias independientes con distribución binomial, con probabilidad de éxito común p, es también binomial. Un resultado similar se obtiene para la suma de dos variables aleatorias Poisson, con se afirma en el siguiente teorema.

Teorema 2.6.11. Sea $X_1, \ldots, X_n$ una sucesión de variables aleatorias independientes de una misma población discreta. Considere $0 < p < 1$.

1. Si $X_i \sim \text{bin}(k_i, p)$, con $k_1, \ldots, k_n \geq 1$, entonces $X_1 + \cdots + X_n \sim \text{bin}(k_1 + \cdots + k_n, p)$.

2. Si $X_i \sim \text{binneg}(r_i, p)$, con $r_1, \ldots, r_n \geq 1$, entonces $X_1 + \cdots + X_n \sim \text{binneg}(r_1 + \cdots + r_n, p)$.

3. Si $X_i \sim \text{Po}(\mu_i)$, con $\mu_1, \ldots, \mu_n > 0$, entonces $X_1 + \cdots + X_n \sim \text{Po}(\mu_1 + \cdots + \mu_n)$.

Demostración. Para la demostración, se puede aplicar tanto la función generadora de momentos como la función generadora de probabilidad. Sea $X = X_1 + \cdots X_n$. Por el Teorema 2.6.10.3, la independencia de estas variables aleatorias implica que la FGP de X sea el producto de las respectivas FGP de X_i:

$$\phi_X(t) = \mathrm{E}\, t^X = \mathrm{E}\, t^{X_1 + \cdots + X_n} = \mathrm{E}\, t^{X_1} \cdots \mathrm{E}\, t^{X_n}; \quad t \geq 0.$$

1. Como $X_i \sim \text{bin}(k_i, p)$, entonces

$$\phi_X(t) = \mathrm{E}\, t^{X_1} \cdots \mathrm{E}\, t^{X_n} = (p(t-1)+1)^{k_1} \cdots (p(t-1)+1)^{k_n} = (p(t-1)+1)^{k_1 + \cdots + k_n}.$$

Esta expresión corresponde a la FGP de la distribución binomial de parámetros $k_1 + \cdots + k_n$ y p.
2. De manera similar, si $X_i \sim \text{binneg}(r_i, p)$, del Ejercicio 2.11.17, se tiene

$$\phi_{X_i} = \mathrm{E}\, t^{X_i} = \left(\frac{pt}{1-(1-p)t}\right)^{r_i}; \quad \text{para} \quad t < \frac{1}{1-p}.$$

Así

$$\phi_X(t) = \mathrm{E}\, t^{X_1} \cdots \mathrm{E}\, t^{X_n} = \left(\frac{pt}{1-(1-p)t}\right)^{r_1} \cdots \left(\frac{pt}{1-(1-p)t}\right)^{r_n} = \left(\frac{pt}{1-(1-p)t}\right)^{r_1 + \cdots + r_n}.$$

3. Por último, si $X_i \sim \text{Po}(\mu_i)$, entonces

$$\phi_X(t) = \mathrm{E}\, t^{X_1} \cdots \mathrm{E}\, t^{X_n} = e^{\mu_1(t-1)} \cdots e^{\mu_n(t-1)} = e^{(\mu_1 + \cdots \mu_n)(t-1)}$$

□

Como consecuencia de este teorema, se confirma que la suma de n variables aleatorias independientes de distribución Bernoulli con parámetro p, es una variable aleatoria de distribución binomial. De hecho, si

$$X \sim \text{bin}(n, p) \quad \text{y} \quad X_1, \ldots, X_n \sim \text{i.i.d.}\, \text{Ber}(p),$$

entonces $X_i \sim \text{bin}(n = 1, p)$; para $i = 1, \ldots, n$ y

$$X \stackrel{d}{=} X_1 + \cdots + X_n.$$

El símbolo $\stackrel{d}{=}$ significa *igualdad en distribución*.

2.7. Aproximación de Poisson

Algunas distribuciones de probabilidad discretas se aproximan a la distribución de Poisson. En tal caso, la distribución de Poisson facilita el cálculo de probabilidades asociadas a la variable original, pues tiene un sólo parámetro. Por ejemplo, si X es una variable aleatoria con distribución binomial $X \sim \text{bin}(n, p)$, con n grande y p pequeño, entonces X tiene una distribución aproximada de Poisson de media

$$\mu = np.$$

De esta manera

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\mu}\frac{\mu^x}{x!}; \quad x = 0, \ldots, n.$$

Se denota como

$$X \overset{\text{aprox}}{\sim} \text{Po}(\mu = np).$$

Por ejemplo, considere un avión con capacidad de 100 pasajeros. Asuma que el 2.5 % de los pasajeros no se presentan a la salida del vuelo. Por lo cual, la aerolínea vende $n = 103$ boletos. Sea la variable aleatoria

$$X = \text{número de pasajeros que no se presentan a la salida del vuelo.}$$

Entonces

$$\mu = np = 103(0.025) = 2.575\,\text{pasajeros}$$

y

$$X \sim \text{bin}(n = 103, p = 0.025) \overset{\text{aprox}}{\sim} \text{Po}(\mu = 2.575).$$

Se compararán las probabilidades exacta y aproximada de los siguientes eventos.

$$\begin{aligned}
A_2 &: \text{ quedaron pasajeros en tierra} = \{X < 3\} = \{X \leq 2\},\\
A_2^c &: \text{ no quedaron pasajeros en tierra} = \{X < 3\}^c = \{X > 2\},\\
A_3 &: \text{ avión parte con cupo lleno} = \{X \leq 3\},\\
A_3^c &: \text{ avión parte con asientos vacíos} = \{X \leq 3\}^c = \{X > 3\}.
\end{aligned}$$

Entonces

$$\begin{aligned}
P(A_2) &= F(2) = P(X \leq 2)\\
&= P(X = 0) + P(X = 1) + P(X = 2)\\
&= f(0) + f(1) + f(2)\\
&= \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} + \binom{n}{2} p^2 (1-p)^{n-2}\\
&= \binom{103}{0} 0.975^{103} + \binom{103}{1} (0.025) 0.975^{102} + \binom{103}{2} 0.025^2 0.975^{101}\\
&= 0.975^{103} + 103(0.025)0.975^{102} + (5253)0.025^2 0.975^{101}\\
&= 0.0737 + 0.1946 + 0.2545\\
&= 0.5229 = 52.29\,\%,
\end{aligned}$$

$$
\begin{aligned}
P(A_3) &= F(3) = P(X \leq 3) \\
&= f(0) + f(1) + f(2) + f(3) \\
&= 0.5229 + \binom{n}{3} p^3 (1-p)^{n-3} = 0.5229 + \binom{103}{3} 0.025^3 0.975^{100} \\
&= 0.5229 + (176851) 0.025^3 0.975^{100} = 0.5229 + 0.2197 \\
&= 0.7426 = 74.26\,\%
\end{aligned}
$$

y

$$
\begin{aligned}
P(A_2^c) &= 1 - P(A_2) = 1 - 0.5229 = 0.4771 = 47.71\,\%, \\
P(A_3^c) &= 1 - P(A_3) = 1 - 0.7426 = 0.2574 = 25.74\,\%.
\end{aligned}
$$

Por aproximación de Poisson, las anteriores probabilidades resultan

$$
\begin{aligned}
P(X \leq 2) \approx e^{-\mu} \left(\frac{\mu^0}{0!} + \frac{\mu^1}{1!} + \frac{\mu^2}{2!} \right) &= e^{-2.575} \left(1 + 2.575 + \frac{2.575^2}{2} \right) \\
&= 0.07615(1 + 2.575 + 3.3153) = 0.5247,
\end{aligned}
$$

$$
\begin{aligned}
P(X \leq 3) \approx e^{-\mu} \left(\frac{\mu^0}{0!} + \frac{\mu^1}{1!} + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} \right) &= 0.5247 + e^{-2.575} \frac{2.575^3}{6} \\
&= 0.5247 + 0.07615(2.8456) = 0.7414
\end{aligned}
$$

y

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) = 1 - 0.5247 = 0.4753, \\
P(X > 3) &= 1 - P(X \leq 3) = 1 - 0.7414 = 0.2586.
\end{aligned}
$$

El error de la aproximación es menor que 0.002. La calidad de la aproximación de Poisson se aprecia en la Figura 2.14, donde se muestran las gráficas de las funciones de densidad y de distribución $\text{bin}(n = 103, p = 0.025)$, junto con las respectivas aproximaciones de la distribución $\text{Po}(\mu = 2.575)$. El siguiente teorema formaliza la aproximación de Poisson de la distribución binomial.

Teorema 2.7.1. Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias de distribución binomial

$$
X_n \sim \text{bin}(n, p_n); \quad n \geq 1,
$$

tal que

$$
\lim_{n \to \infty} n p_n = \mu > 0.
$$

Entonces

$$
\lim_{n \to \infty} P(X_n = x) = e^{-\mu} \frac{\mu^x}{x!}; \quad x \geq 0. \tag{2.89}
$$

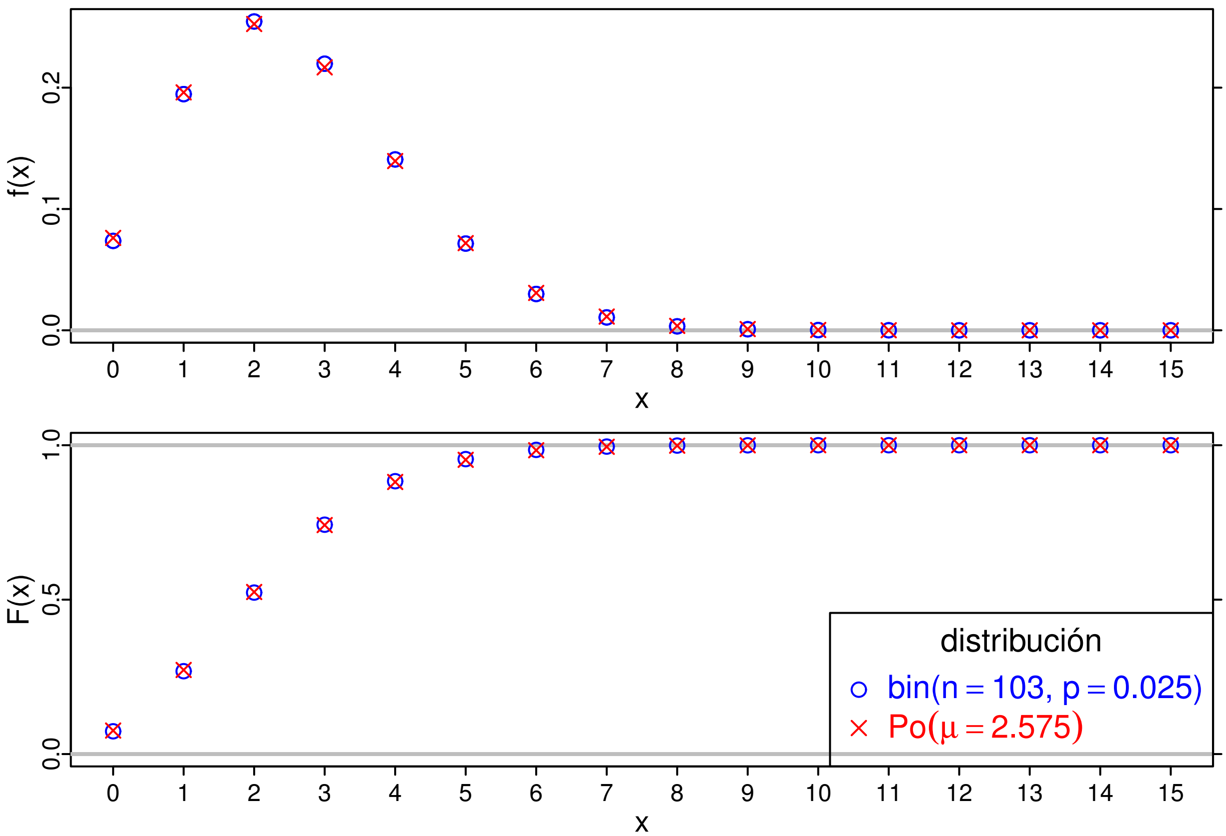


Figura 2.14. [arriba] Gráfica de funciones de densidad binomial y de aproximación de Poisson. [abajo] Respectivas funciones de distrbución.

Demostración. Para $x \geq 0$, se tiene

$$\begin{aligned}
P(X_n = x) &= \binom{n}{x} p_n^x (1-p_n)^{n-x} \\
&= \frac{n!}{x!(n-x)!} \frac{(np_n)^x}{n^x} \frac{(1-np_n/n)^n}{(1-p_n)^x} \\
&= \frac{1}{x!} \frac{n!}{(n-x)!n^x} \left(\frac{np_n}{1-p_n}\right)^x \left(1-\frac{np_n}{n}\right)^n .
\end{aligned}$$

Se calculará el límite del miembro derecho de la expresión anterior. Note que

$$\begin{aligned}
\lim_{n\to\infty} p_n &= \lim_{n\to\infty} \frac{np_n}{n} = \frac{\mu}{\infty} = 0, \\
\lim_{n\to\infty} \left(\frac{np_n}{1-p_n}\right)^x &= \left[\lim_{n\to\infty} \frac{np_n}{1-p_n}\right]^x = \left(\frac{\mu}{1}\right)^x = \mu^x, \\
\lim_{n\to\infty} \left(1-\frac{np_n}{n}\right)^n &= e^{-\mu}.
\end{aligned}$$

El último límite se obtiene por (21) del Apéndice. Por último, para $x \geq 1$, se tiene

$$\begin{aligned}
\lim_{n\to\infty} \frac{n!}{(n-x)!n^x} &= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n\times n\times\cdots\times n} \\
&= \lim_{n\to\infty} \left[\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right] \\
&= \lim_{n\to\infty} \frac{n}{n} \lim_{n\to\infty} \frac{n-1}{n} \cdots \lim_{n\to\infty} \frac{n-x+1}{n} \\
&= 1^x = 1.
\end{aligned}$$

Se concluye (2.89). $\square$

2.8. Distribución condicional

Sea (X, Y) un vector aleatorio discreto con función de densidad conjunta $f(x, y)$ (2.40). La función de *densidad condicional* de Y dado el evento $\{X = x\}$ es

$$f(y \mid x) = P(Y = y \mid X = x); \quad x, y = 0, 1, \ldots,$$

siempre que $f(x) > 0$, y cero en otros casos. Como

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{f(x)},$$

entonces

$$f(y \mid x) = \frac{f(x, y)}{f(x)}; \quad x, y = 0, 1, \ldots$$

De la misma manera, dado el evento $\{Y = y\}$, la función de *densidad condicional* de la variable aleatoria X es

$$f(x \mid y) = P(X = x \mid Y = y) = \frac{f(x,y)}{f(y)}; \quad x, y = 0, 1, \ldots,$$

siempre que $f(y) > 0$, y cero en otros casos. Así, la función de densidad conjunta de X y Y se escribe de dos maneras diferentes:

$$f(x,y) = f(x)f(y \mid x) = f(x \mid y)f(y); \quad \text{para} \quad x, y = 0, 1, \ldots$$

En particular, las variables aleatorias X y Y son independientes, si y sólo si se cumple cualesquiera de las siguientes afirmaciones

$$\begin{aligned} f(y \mid x) &= f(y); \quad \text{para todo} \quad x, y = 0, 1, \ldots, \quad \text{con} \quad f(x) > 0, \\ f(x \mid y) &= f(x); \quad \text{para todo} \quad x, y = 0, 1, \ldots, \quad \text{con} \quad f(y) > 0. \end{aligned}$$

De hecho, X y Y son variables aleatorias independientes si y sólo si

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{f(x)f(y)}{f(x)} = f(y) \quad x, y = 0, 1, \ldots$$

La otra equivalencia se verifica de manera similar.

En el caso de *dependencia*, la distribución de la variable aleatoria Y es sensible al valor dado $\{X = x\}$. Al condicionar, la probabilidad de los eventos $\{Y = y\}$ se reajustan. Según el contexto, aplican dos funciones de densidad para la variable aleatoria Y. La probabilidad $f(y)$ aplica cuando no se tiene información de la variable aleatoria X. En cambio, la capacidad predictiva de la variable aleatoria Y se mejora cuando se conoce de antemano el evento $\{X = x\}$. Por lo que aplica la probabilidad condicional $f(y \mid x)$.

Dado el evento $\{X = x\}$, la distribución de probabilidad condicional de la variable aleatoria Y tiene asociados todos los conceptos poblacionales análogos a su contraparte incondicional. En particular, su *esperanza condicional* es

$$\mathrm{E}[Y \mid X = x] = \sum_{y=0}^{\infty} y f(y \mid x); \quad \text{para} \quad x = 0, 1, \ldots \tag{2.90}$$

Análogo a (2.26), se tiene

$$\mathrm{E}[g(Y) \mid X = x] = \sum_{y=0}^{\infty} g(y) f(y \mid x); \quad \text{para} \quad x = 0, 1, \ldots, \tag{2.91}$$

donde $g(y)$ es una función real. En particular, se tiene

$$\mathrm{E}[Y^2 \mid X = x] = \sum_{y=0}^{\infty} y^2 f(y \mid x)$$

mientras que la *varianza condicional* es

$$
\begin{aligned}
\mathrm{V}[Y \mid X = x] &= \mathrm{E}[(Y - \mathrm{E}[Y \mid X = x])^2 \mid X = x] \\
&= \mathrm{E}[Y^2 \mid X = x] - (\mathrm{E}[Y \mid X = x])^2. \qquad (2.92)
\end{aligned}
$$

La última igualdad se debe al Ejercicio 2.21.

La expresión (2.90) es una función real dependiente de la variable entera x. Si se evalúa en $x = X$, la esperanza condicional adquiere estatus de variable aleatoria:

$$
\mathrm{E}[Y \mid X] = \sum_{y=0}^{\infty} y f(y \mid X).
$$

Su conjunto de valores posibles es

$$
\mathrm{E}[Y \mid X] = \mathrm{E}[Y \mid X = 0],\ \mathrm{E}[Y \mid X = 1],\ \mathrm{E}[Y \mid X = 2], \ldots
$$

Así mismo, la varianza condicional (2.92) se escribe como

$$
\mathrm{V}[Y \mid X] = \mathrm{E}[(Y - \mathrm{E}[Y \mid X])^2 \mid X] = \mathrm{E}[Y^2 \mid X] - (\mathrm{E}[Y \mid X])^2.
$$

El siguiente teorema resume las propiedades básicas de la esperanza condicional, muchas de las cuales son análogas a las de su contraparte incondicional.

Teorema 2.8.1. Sean X, Y, Y_1, Y_2 variables aleatorias con esperanza finita (no necesariamente discretas). Considere la constante $b \in \mathbf{R}$ y las funciones reales $g(x)$, $h(y)$ y $g(x, y)$. Entonces

1. $$\mathrm{E}[Y_1 + bY_2 \mid X] = \mathrm{E}[Y_1 \mid X] + b\,\mathrm{E}[Y_2 \mid X].$$

2. Si $Y_1 \leq Y_2$, entonces
$$\mathrm{E}[Y_1 \mid X] \leq \mathrm{E}[Y_2 \mid X].$$
En particular, $\mathrm{E}[Y \mid X] \geq 0$, siempre que $Y \geq 0$.

3. Si X y Y son variables aleatorias independientes, entonces
$$\mathrm{E}[Y \mid X] = \mathrm{E}\,Y.$$

4. $$\mathrm{E}[g(X, Y) \mid X = x] = \mathrm{E}[g(x, Y) \mid X = x].$$
En particular
$$
\begin{aligned}
\mathrm{E}[g(X) \mid X] &= g(X), \qquad (2.93) \\
\mathrm{E}[g(X)h(Y) \mid X] &= g(X)\,\mathrm{E}[h(Y) \mid X], \\
\mathrm{E}[\mathrm{E}[g(X) \mid X] \mid X] &= \mathrm{E}[g(X) \mid X].
\end{aligned}
$$

Como cualquier otra variable aleatoria con esperanza finita, $\mathrm{E}[Y \mid X]$ tiene media poblacional, la cual coincide con su contraparte incondicional:

$$\mathrm{E}\{\mathrm{E}[Y \mid X]\} = \mathrm{E}\,Y. \tag{2.94}$$

Esta es la ley de *esperanza iterada*, que se deduce por (2.26) y (2.90):

$$\begin{aligned}
\mathrm{E}\{\mathrm{E}[Y \mid X]\} &= \sum_{x=0}^{\infty} \mathrm{E}[Y \mid X = x] f(x) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y f(y \mid x) f(x) \\
&= \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} y f(x,y) = \sum_{y=0}^{\infty} y \sum_{x=0}^{\infty} f(x,y) \\
&= \sum_{y=0}^{\infty} y f(y) = \mathrm{E}\,Y.
\end{aligned}$$

Por otro lado, del Ejercicio 2.11.22, se cumple

$$\mathrm{V}\,Y = \mathrm{E}\{\mathrm{V}[Y \mid X]\} + \mathrm{V}\{\mathrm{E}[Y \mid X]\}. \tag{2.95}$$

Así que la variable aleatoria $\mathrm{E}[Y \mid X]$ tiene la misma media de Y, pero con una menor varianza.

Por el teorema 2.8.1, se tiene lo siguiente. La propiedad 2.8.1.1 significa que la esperanza condicional es un operador lineal. Ante la independencia de X y Y, la propiedad 2.8.1.3 se debe que la variable aleatoria X no aporta información relevante de Y. Esta propiedad que contrasta con 2.8.1.2.93, donde la información de X sobre $g(X)$ es total. La propiedad 2.8.1.4 significa que, dado $\{X = x\}$, el argumento X debe verse como una constante en la expresión funcional $g(X,Y)$.

Ejemplo 2.8.2. Sean X y Y dos variables aleatorias independientes con distribución geométrica de parámetro p, con conjunto de valores posibles $x, y = 0, 1, 2, \ldots$ Recuerde que ambas variables aleatorias cuentan el número de fracasos hasta el respectivo primer éxito. Encuentre:

1. La distribución de $Z = X + Y$.

2. La distribución condicional de X dado $Z = z$.

Solución. 1. La función de densidades marginal, tanto de X como Y, es

$$f_X(x) = f_Y(x) = P(X = x) = P(Y = x) = p(1-p)^x; \quad x = 0, 1, \ldots$$

Por la independencia de las variables aleatorias X y Y, la función de densidad conjunta es

$$f(x,y) = f(x)f(y) = p(1-p)^x \cdot p(1-p)^y = p^2(1-p)^{x+y}; \quad x, y = 0, 1, \ldots$$

El conjunto de valores posibles de la variable aleatoria Z es $\{0, 1, 2, \ldots\}$. Note que

$$\{Z = z\} = \{X + Y = z\} = \bigcup_{x=0}^{z} \{X = x, Y = z - x\}. \tag{2.96}$$

Esta es una unión de eventos excluyentes. La Figura 2.15 representa el conjunto de vectores posibles de (X, Y), junto con los vectores posibles del evento $\{Z = 5\}$. Así

$$\begin{aligned}
P(Z = z) &= P\left(\bigcup_{x=0}^{z}\{X = x, Y = z - x\}\right) = \sum_{x=0}^{z} P(X = x, Y = z - x) \\
&= \sum_{x=0}^{z} f_{X,Y}(x, z - x) = \sum_{x=0}^{z} p^2(1-p)^{x+(z-x)} \\
&= p^2 \sum_{x=0}^{z}(1-p)^z = p^2(1-p)^z(z+1); \quad z = 0, 1, \ldots
\end{aligned}$$

2. Dado el evento $\{Z = z\}$, el conjunto de valores posibles de X se reduce a $0, 1, \ldots, z$. Así que

$$\begin{aligned}
P(X = x \mid Z = z) &= \frac{P(X = x, Z = z)}{P(Z = z)} = \frac{P(X = x, X + Y = z)}{P(Z = z)} \\
&= \frac{P(X = x, x + Y = z)}{f_Z(z)} = \frac{P(X = x, Y = z - x)}{f_Z(z)} \\
&= \frac{f_{X,Y}(x, z - x)}{f_Z(z)} = \frac{p^2(1-p)^{x+(z-x)}}{p^2(1-p)^z(z+1)} = \frac{1}{z+1}.
\end{aligned}$$

Se concluye que la distribución condicional de la variable aleatoria X, dado $Z = z$, es uniforme discreta en $\{0, 1, \ldots, z\}$:

$$[X \mid Z = z] \sim \mathrm{U}(0, 1, \ldots, z).$$

Compare la distribución condicional de X con su contraparte marginal. Por el Ejercicio 2.8, se tiene

$$\mathrm{E}[X \mid Z = z] = \frac{0 + z}{2} = \frac{z}{2}$$

y

$$\mathrm{V}[X \mid Z = z] = \frac{(z - 0 + 1)^2 - 1}{12} = \frac{z(z+2)}{12}.$$

Se comprueba también la fórmula de la esperanza iterada (2.94):

$$\mathrm{E}\{\mathrm{E}[X \mid Z]\} = \mathrm{E}\left[\frac{Z}{2}\right] = \mathrm{E}\left[\frac{X + Y}{2}\right] = \frac{1}{2} \cdot 2\,\mathrm{E}\,X = \mathrm{E}\,X.$$

El segundo momento condicional satisface

$$\begin{aligned}
\mathrm{E}[X^2 \mid Z = z] &= \mathrm{V}[X \mid Z = z] + (\mathrm{E}[X \mid Z = z])^2 \\
&= \frac{z(z+2)}{12} + \left(\frac{z}{2}\right)^2 = \frac{z(z+2)}{12} + \frac{z^2}{4} \\
&= \frac{z^2 + 2z + 3z^2}{12} = \frac{z(1 + 2z)}{6}.
\end{aligned}$$

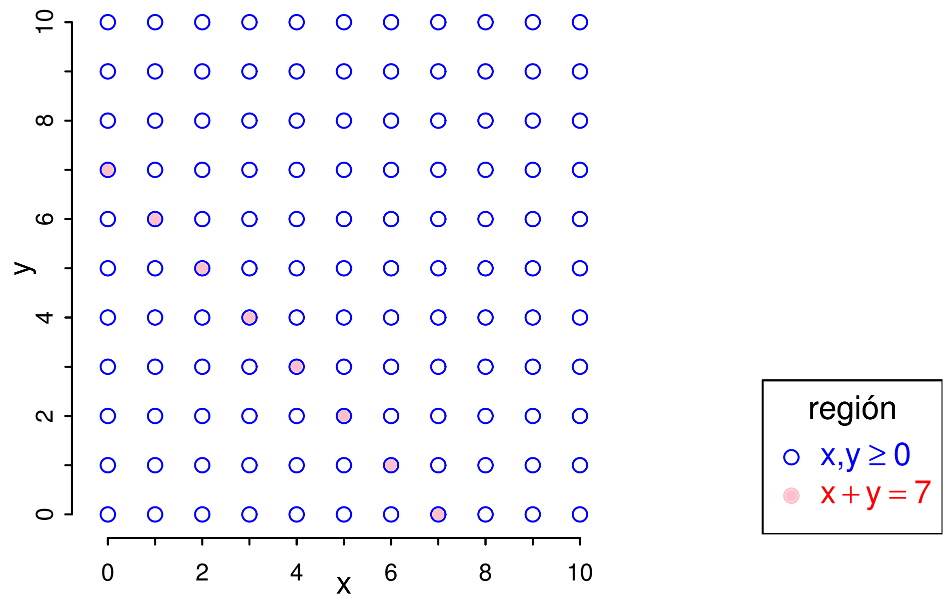


Figura 2.15. Representación gráfica del conjunto de vectores posibles de (X, Y), así como del evento $\{X + Y = z\}$, con $z = 7$; Ejemplo 2.8.2.

Así mismo, se satisface (2.95):

$$\begin{aligned}
&\mathrm{E}\{\mathrm{V}[X \mid Z]\} + \mathrm{V}\{\mathrm{E}[X \mid Z]\} = \mathrm{E}\left\{\frac{Z(Z+2)}{12}\right\} + \mathrm{V}\left\{\frac{Z}{2}\right\} \\
&= \frac{\mathrm{E}\,Z^2 + 2\,\mathrm{E}\,Z}{12} + \frac{\mathrm{V}\,Z}{4} = \frac{[\mathrm{V}\,Z + (\mathrm{E}\,Z)^2] + 2\,\mathrm{E}\,Z + 3\,\mathrm{V}\,Z}{12} \\
&= \frac{4\,\mathrm{V}[X+Y] + (\mathrm{E}[X+Y])^2 + 2\,\mathrm{E}[X+Y]}{12} \\
&= \frac{4 \cdot 2(1-p)/p^2 + (2(1-p)/p)^2 + 2 \cdot 2(1-p)/p}{12} \\
&= (1-p)\frac{2 + (1-p) + p}{3p^2} = \frac{1-p}{p^2} \\
&= \mathrm{V}\,X.
\end{aligned}$$

Ejemplo 2.8.3. Un programa piloto de combate de la pobreza extrema consiste en dar un apoyo económico a cada familia que tenga un estudiante regular menor de edad, en una comunidad de $m = 500$ familias. Sin embargo, por causa de la migración forzada, la comunidad redujo su población a N familias, donde N se modela como una distribución binomial:

$$N \sim \mathrm{bin}(m = 500, q).$$

El parámetro q es la probabilidad de cada familia candidata de permanecer en el pueblo durante 5 años. Por otro lado, diversas causas de logística ocasiona que el programa tenga una cobertura de $100p\,\%$ de las N familias. Se desea conocer la distribución del total de familias X que finalmente se implementó el programa. Note que la distribución condicional de la variable aleatoria X, dado N, es binomial de parámetros N y q. Además N tiene distribución binomial de parámetros $m = 500$ y p:

$$[X \mid N] \sim \text{bin}(N, p) \quad \text{y} \quad N \sim \text{bin}(m = 500, q).$$

Obtenga las funciones de densidad conjunta de (X, Y) y marginal de X. Primero, identifique los vectores posibles de (X, N). Como $0 \leq X \leq N \leq m$, entonces el conjunto de vectores posibles de (X, N) es

$$\{(x, y) : 0 \leq x \leq y \leq m\}.$$

La función de densidad conjunta de (X, N) es

$$\begin{aligned}
f(x, y) &= P(X = x, N = y) = P(X = x \mid N = y)P(N = y) \\
&= \binom{y}{x} p^x (1-p)^{y-x} \cdot \binom{m}{y} q^y (1-q)^{m-y} \\
&= \frac{y!}{x!(y-x)!} \frac{m!}{y!(m-y)!} p^x (1-p)^{y-x} q^y (1-q)^{m-y} \\
&= \frac{m!}{x!(m-x)!} \frac{(m-x)!}{(y-x)!(m-y)!} p^x (1-p)^{y-x} q^y (1-q)^{m-y} \\
&= \binom{m}{x} \binom{m-x}{y-x} p^x (1-p)^{y-x} q^y (1-q)^{m-y}; \quad 0 \leq x \leq y \leq m.
\end{aligned}$$

La función de densidad marginal de X se obtiene como sigue. Incondicionalmente, X toma los valores

$$X = 0, \ldots, m.$$

Así, para cualquier $0 \leq x \leq m$, se tiene

$$\begin{aligned}
f(x) &= \sum_{y=0}^{\infty} f(x, y) = \sum_{y=x}^{m} f(x, y) \\
&= \sum_{y=x}^{m} \binom{m}{x} \binom{m-x}{y-x} p^x (1-p)^{y-x} q^y (1-q)^{m-y} \\
&= \binom{m}{x} p^x \sum_{y=x}^{m} \binom{m-x}{y-x} (1-p)^{y-x} q^{(y-x)+x} (1-q)^{(m-x)-(y-x)} \\
&= \binom{m}{x} p^x q^x (1-q)^{m-x} \sum_{y=x}^{m} \binom{m-x}{y-x} (1-p)^{y-x} q^{y-x} (1-q)^{-(y-x)} \\
&= \binom{m}{x} (pq)^x (1-q)^{m-x} \sum_{k=0}^{m-x} \binom{m-x}{k} \left(\frac{(1-p)q}{1-q} \right)^k \\
&= \binom{m}{x} (pq)^x (1-q)^{m-x} \left[1 + \frac{(1-p)q}{1-q} \right]^{m-x} \\
&= \binom{m}{x} (pq)^x (1-pq)^{m-x}.
\end{aligned}$$

Se concluye

$$X \sim \mathrm{bin}(m, pq),$$

siempre que

$$[X \mid N] \sim \mathrm{bin}(N, p) \quad \text{y} \quad N \sim \mathrm{bin}(m, q).$$

Ejemplo 2.8.4. Considere una variación del ejemplo anterior, donde ahora N es una variable aleatoria de distribución Poisson de media $\lambda > 0$. Entonces

$$[X \mid N] \sim \mathrm{bin}(N, p) \quad \text{y} \quad N \sim \mathrm{Po}(\lambda).$$

Se verificará que

$$X \sim \mathrm{Po}(\lambda p).$$

La función de densidad conjunta de (X, N) es

$$\begin{aligned} f(x,y) &= P(X = x, N = y) = P(X = x \mid N = y)P(N = y) \\ &= \binom{y}{x} p^x (1-p)^{y-x} \cdot e^{-\lambda}\frac{\lambda^y}{y!} = \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} e^{-\lambda}\frac{\lambda^y}{y!} \\ &= \frac{e^{-\lambda}}{x!(y-x)!} p^x (1-p)^{y-x} \lambda^y; \quad \text{para} \quad 0 \le x \le y. \end{aligned}$$

Así, la función de densidad marginal de X es

$$\begin{aligned} f(x) &= \sum_{y=0}^{\infty} f(x,y) = \sum_{y=x}^{\infty} f(x,y) \\ &= \sum_{y=x}^{\infty} \frac{e^{-\lambda}}{x!(y-x)!} p^x (1-p)^{y-x} \lambda^y = \frac{e^{-\lambda}}{x!} p^x \sum_{y=x}^{\infty} \frac{1}{(y-x)!} (1-p)^{y-x} \lambda^{(y-x)+x} \\ &= \frac{e^{-\lambda}}{x!} \lambda^x p^x \sum_{y=x}^{\infty} \frac{1}{(y-x)!} [\lambda(1-p)]^{y-x} = \frac{e^{-\lambda}}{x!} (\lambda p)^x \sum_{z=0}^{\infty} \frac{1}{z!} [\lambda(1-p)]^z \\ &= \frac{e^{-\lambda}}{x!} (\lambda p)^x e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^x}{x!}; \quad \text{para} \quad x \ge 0. \end{aligned}$$

Por lo tanto, X tiene una distribución de Poisson de media λp. Compare las medias incondicional y condicional de X, así como sus respectivas varianzas:

$$\begin{aligned} \mathrm{E}\, X = \lambda p \quad &\text{y} \quad \mathrm{E}[X \mid N] = Np, \\ \mathrm{V}\, X = \lambda p \quad &\text{y} \quad \mathrm{V}[X \mid N] = Np(1-p). \end{aligned}$$

Además

$$\mathrm{E}\{\mathrm{E}[X \mid N]\} = \mathrm{E}\{Np\} = p\,\mathrm{E}\, N = \lambda p = \mathrm{E}\, X$$

y

$$\mathrm{E}\{\mathrm{V}[X \mid N]\} + \mathrm{V}\{\mathrm{E}[X \mid N]\} = \mathrm{E}\{Np(1-p)\} + \mathrm{V}\{Np\} = \lambda p(1-p) + \lambda p^2 = \lambda p = \mathrm{V}\, X.$$

Ejemplo 2.8.5. Sean X y Y dos variables aleatorias independientes con distribución geométrica de parámetro $0 < p < 1$; con valores posibles en $0, 1, \ldots$

1. Calcule la probabilidad del evento $\{X \wedge Y = X\} = \{X \leq Y\}$.
2. Encuentre la distribución del mínimo $X \wedge Y$.
3. Encuentre la distribución de la suma $X + Y$.

Solución. 1. La función de densidad de las variables aleatorias X y Y está dada por (2.21). Sean los eventos $A = \{X \leq Y\}$ y $B = \{X \geq Y\}$. El rol de ambas variables aleatorias es intercambiable, pues son independientes e idénticamente distribuidas. Entonces, dichos eventos son igualmente probables. Además

$$A \cup B = \Omega \quad \text{y} \quad A \cap B = \{X = Y\}.$$

Así

$$1 = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2P(A) - P(X = Y),$$

con

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

Por lo que

$$P(A) = \frac{1 + P(X = Y)}{2} = \frac{1 + p/(2-p)}{2} = \frac{(2-p)+p}{2(2-p)} = \frac{1}{2-p}.$$

2. Sea $W = X \wedge Y$. Sus valores posibles son

$$W = 0, 1, \ldots$$

Para $w = 0, 1, \ldots$, se tiene

$$\{W > w\} = \{X \wedge Y > w\} = \{X > w, Y > w\} = \{X > w\} \cap \{Y > w\}.$$

Esta es una intersección de eventos independientes y equiprobables. Por lo que

$$\begin{aligned} P(W > w) &= P(X > w, Y > w) = P(X > w)P(Y > w) \\ &= (1-p)^{w+1}(1-p)^{w+1} = (1-p)^{2(w+1)}. \end{aligned}$$

Aquí se usó el Ejercicio 2.11.5. Así

$$P(W \le w) = 1 - P(W > w) = 1 - (1-p)^{2(w+1)}; \quad w = 0, 1, \ldots$$

La función de densidad de W es

$$\begin{aligned} P(W = w) &= P(W \le w) - P(W \le w-1) = F(w) - F(w-1) \\ &= [1-(1-p)^{2(w+1)}] - [1-(1-p)^{2w}] = (1-p)^{2w} - (1-p)^{2w+2} \\ &= (1-p)^{2w}[1-(1-p)^2] = p(2-p)(1-p)^{2w}. \end{aligned}$$

Note que, al comparar esta expresión con (2.21), la distribución de la variable aleatoria W es geométrica con parámetro $p(2-p) \in (p, 1)$. Esta distribución de probabilidad está más concentrada cerca de cero que la de las variable aleatorias X y Y.
3. Sea la variable aleatoria $Z = X + Y$, cuyo conjunto de valores posibles es $Z = 0, 1, \ldots$ Al considerar (2.96), se tiene

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^{z} P(X = x, Y = z - x) = \sum_{x=0}^{z} P(X = x) P(Y = z - x) \\ &= \sum_{x=0}^{z} p(1-p)^x \cdot p(1-p)^{z-x} = p^2 \sum_{x=0}^{z} (1-p)^z \\ &= p^2(z+1)(1-p)^z; \quad \text{para} \quad z = 0, 1, \ldots. \end{aligned}$$

Estas probabilidades suman uno:

$$\begin{aligned} \sum_{z=0}^{\infty} P(Z = z) &= \sum_{z=0}^{\infty} p^2(z+1)(1-p)^z = p^2 \sum_{z=0}^{\infty} z(1-p)^z + p^2 \sum_{z=0}^{\infty} (1-p)^z \\ &= p \sum_{z=0}^{\infty} zp(1-p)^z + p \sum_{z=0}^{\infty} p(1-p)^z = p(\mathrm{E}\, X + 1) \\ &= p\left(\frac{1-p}{p} + 1\right) = 1. \end{aligned}$$

Ejemplo 2.8.6. Considere la función de densidad conjunta

$$f(x, y) = e^{-(\lambda+\mu)} \frac{\lambda^x \mu^y}{x! y!}; \quad x, y = 0, 1, \ldots,$$

con $\lambda, \mu > 0$. Note que $f(x, y)$ es el producto de dos funciones de densidad univariadas, ambas de distribución de Poisson:

$$f(x, y) = e^{-(\lambda+\mu)} \frac{\lambda^x \mu^y}{x! y!} = e^{-\lambda} \frac{\lambda^x}{x!} \cdot e^{-\mu} \frac{\mu^y}{y!} = f(x) f(y); \quad x, y = 0, 1, \ldots.$$

Por lo que X y Y son variables aleatorias independientes, con $X \sim \mathrm{Po}(\lambda)$ y $Y \sim \mathrm{Po}(\mu)$.

Además $Z = X + Y \sim \text{Po}(\lambda + \mu)$. De hecho, por (2.96), se tiene

$$\begin{aligned}
P(Z = z) &= \sum_{x=0}^{z} P(X = x, Y = z - x) = \sum_{x=0}^{z} P(X = x)P(Y = z - x) \\
&= \sum_{x=0}^{z} e^{-\lambda} \frac{\lambda^x}{x!} \cdot e^{-\mu} \frac{\mu^{z-x}}{(z-x)!} = e^{-\lambda-\mu} \frac{1}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x} \\
&= e^{-(\lambda+\mu)} \frac{1}{z!} \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^z}{z!}; \quad z = 0, 1, \ldots
\end{aligned}$$

Ejemplo 2.8.7. Considere la función

$$f(x, y) = k \frac{\lambda^x \mu^{y-x}}{x!(y-x)!}; \quad \text{para} \quad 0 \leq x \leq y,$$

con cierta constante $k > 0$.

1. ¿Cuanto vale la constante k para que $f(x, y)$ sea una función de densidad?

2. Obtenga las funciones de densidad marginal de las variables aleatorias X y Y.

Solución. 1. El conjunto de vectores posibles de (X, Y) se describe en la Figura 2.10; punto rosa. Como $P(X \leq Y) = 1$, se descarta que X y Y sean variables aleatorias independientes. Considere

$$\begin{aligned}
1 &= \sum_{x,y=0}^{\infty} f(x, y) = \sum_{0 \leq x \leq y} f(x, y) \\
&= \sum_{y=0}^{\infty} \sum_{x=0}^{y} f(x, y) = \sum_{y=0}^{\infty} \sum_{x=0}^{y} k \frac{\lambda^x \mu^{y-x}}{x!(y-x)!} \\
&= k \sum_{y=0}^{\infty} \frac{1}{y!} \sum_{x=0}^{y} \frac{y!}{x!(y-x)!} \lambda^x \mu^{y-x} = k \sum_{y=0}^{\infty} \frac{1}{y!} \sum_{x=0}^{y} \binom{y}{x} \lambda^x \mu^{y-x} \\
&= k \sum_{y=0}^{\infty} \frac{(\lambda+\mu)^y}{y!} = k e^{\lambda+\mu}.
\end{aligned}$$

Así, $k = e^{-(\lambda+\mu)}$ y la función de densidad conjunta de (X, Y) es

$$f(x, y) = e^{-(\lambda+\mu)} \frac{\lambda^x \mu^{y-x}}{x!(y-x)!}; \quad 0 \leq x \leq y.$$

Por el Teorema de Tonelli Fubini, este resultado coincide al intercambiar las sumatorias:

$$\begin{aligned}
\sum_{0 \leq x \leq y} f(x, y) &= k \sum_{x=0}^{\infty} \sum_{y=x}^{\infty} \frac{\lambda^x \mu^{y-x}}{x!(y-x)!} = k \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \sum_{y=x}^{\infty} \frac{\mu^{y-x}}{(y-x)!} \\
&= k \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \sum_{z=0}^{\infty} \frac{\mu^z}{z!} = k \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{\mu} \\
&= k e^{\mu} e^{\lambda} = k e^{\lambda+\mu}.
\end{aligned}$$

2. Ahora se obtendrán las funciones de densidades marginales de X y Y. Ambas variables aleatorias comparten el conjunto de valores posibles, que son los enteros no negativos. Así

$$\begin{aligned} f(x) = P(X = x) &= \sum_{y=0}^{\infty} f(x,y) \\ &= e^{-(\lambda+\mu)} \sum_{y=x}^{\infty} \frac{\lambda^x \mu^{y-x}}{x!(y-x)!} = e^{-(\lambda+\mu)} \frac{\lambda^x}{x!} \sum_{y=x}^{\infty} \frac{\mu^{y-x}}{(y-x)!} \\ &= e^{-(\lambda+\mu)} \frac{\lambda^x}{x!} \sum_{z=0}^{\infty} \frac{\mu^z}{z!} = e^{-(\lambda+\mu)} \frac{\lambda^x}{x!} e^{\mu} \\ &= e^{-\lambda} \frac{\lambda^x}{x!}; \quad \text{para} \quad x = 0, 1, \ldots \end{aligned}$$

Por lo tanto, X es una variable aleatoria de distribución de Poisson con media λ. De manera similar, la función de densidad marginal de Y es Poisson de con $\lambda + \mu$:

$$\begin{aligned} f(y) = P(Y = y) &= \sum_{x=0}^{\infty} f(x,y) \\ &= \sum_{x=0}^{y} f(x,y) = e^{-(\lambda+\mu)} \sum_{x=0}^{y} \frac{\lambda^x \mu^{y-x}}{x!(y-x)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{y!} \sum_{x=0}^{y} \binom{y}{x} \lambda^x \mu^{y-x} = e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^y}{y!}; \quad \text{para} \quad y = 0, 1, \ldots \end{aligned}$$

Esta sección concluye con el desarrollo de vectores aleatorios multidimensionales, especialmente para el caso discreto.

Un *vector aleatorio* $(X_1, \ldots, X_n)$ en $\mathbf{R}^n$ es una colección de $n \geq 2$ variables aleatorias definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Su función de *distribución conjunta* es

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n); \quad \text{para} \quad x_1, \ldots, x_n \in \mathbf{R}.$$

El vector aleatorio $(X_1, \ldots, X_n)$ es discreto si sus elementos son variables aleatorias discretas. En tal caso, se asume que $X_1, \ldots, X_n$ toman valores en los enteros no negativos: $x_1, \ldots, x_n = 0, 1, 2, \ldots$ La función de *densidad conjunta* de $(X_1, \ldots, X_n)$ se define como

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n); \quad \text{para} \quad x_1, \ldots, x_n = 0, 1, \ldots$$

Se verifica que la función de *densidad marginal* de X_i; para $i = 1, \ldots, n$, es la suma de la función de densidad conjunta, sobre todos los valores posibles del resto de las $n-1$ variables aleatorias:

$$f(x_i) = P(X_i = x_i) = \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n = 0, 1, \ldots} f(x_1, \ldots, x_n); \quad x_i = 0, 1, \ldots$$

En particular, la función de densidad marginal de X_1 es

$$f(x_1) = P(X_1 = x_1) = \sum_{x_2, \ldots, x_n = 0, 1, \ldots} f(x_1, \ldots, x_n); \quad x_1 = 0, 1, \ldots$$

Considere dos conjuntos de índices, todos diferentes entre si: $1 \leq i_1 < \cdots < i_k \leq n$ y $1 \leq j_1 < \cdots < j_l \leq n$, con $2 \leq k + l \leq n$. Dado el vector $(X_{i_1}, \ldots, X_{i_k}) = (x_{i_1}, \ldots, x_{i_k})$, con $x_{i_1}, \ldots, x_{i_k} = 0, 1, \ldots$ y $f(x_{i_1}, \ldots, x_{i_k}) > 0$, la función de *densidad condicional* de $(X_{j_1}, \ldots, X_{j_l})$ es

$$f(x_{j_1}, \ldots, x_{j_l} \mid x_{i_1}, \ldots, x_{i_k}) = \frac{f(x_1, \ldots, x_n)}{f(x_{i_1}, \ldots, x_{i_k})}; \quad x_{j_1}, \ldots, x_{j_l} = 0, 1, \ldots$$

La Tabla 2.7 muestra las diversas funciones de densidad de un vector aleatorio discreto de dimensión $n = 3$: conjunta, marginales y condicionales.

Se dice que las variables aleatorias $X_1, \ldots, X_n$, son *independientes* si su función de distribución conjunta es el producto de las funciones de distribución marginales:

$$F(x_1, \ldots, x_n) = F(x_1) \cdots F(x_n); \quad \text{para todo} \quad x_1, \ldots, x_n \in \mathbf{R}.$$

En el caso de variables aleatorias discretas, este concepto equivale a que la función de densidad conjunta sea es el producto de sus funciones de densidad marginales. El siguiente teorema especifica este resultado.

Teorema 2.8.8. Considere el vector de variables aleatorias $(X_1, \ldots, X_n)$, con función de distribución conjunta $F(x_1, \ldots, x_n)$ y respectivas funciones de distribución marginales $F(x_1)$, $\ldots$, $F(x_n)$. Entonces, son equivalentes las siguientes afirmaciones

1. Son independientes las variables aleatorias $X_1, \ldots, X_n$,

2. $$F(x_1, \ldots, x_n) = F(x_1) \cdots F(x_n); \quad x_1, \ldots, x_n \in \mathbf{R},$$

3. $$P(X_1 < x_1, \ldots, X_n < x_n) = P(X_1 < x_1) \cdots P(X_n < x_n); \quad x_1, \ldots, x_n \in \mathbf{R},$$

4. $$P(X_1 \geq x_1, \ldots, X_n \geq x_n) = P(X_1 \geq x_1) \cdots P(X_n \geq x_n); \quad x_1, \ldots, x_n \in \mathbf{R},$$

5. $$P(X_1 > x_1, \ldots, X_n > x_n) = P(X_1 > x_1) \cdots P(X_n > x_n); \quad x_1, \ldots, x_n \in \mathbf{R}.$$

6. Para cualesquier $A_1, \ldots, A_n \in \mathcal{B}(\mathbf{R})$, son independientes los eventos $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$:

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Si en particular $X_1, \ldots, X_n$ son variables aleatorias discretas, con conjunto de valores posibles en los enteros no negativos, entonces las anteriores afirmaciones son equivalentes a su contraparte evaluadas en $x_1, \ldots, x_n = 0, 1, \ldots$, así como $A_1, \ldots, A_n \subset \{0, 1, \ldots\}$. En tal caso, las anteriores afirmaciones son también equivalentes a

7. $$f(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n); \quad \text{para todo} \quad x_1, \ldots, x_n = 0, 1, \ldots$$

tipo de densidad	variable o vector	función de densidad
conjunta	(X,Y,Z)	$f(x,y,z) = P(X=x, Y=y, Z=z)$
marginal	(X,Y)	$f(x,y) = P(X=x, Y=y) = \sum_{z=0}^{\infty} f(x,y,z)$
marginal	(X,Z)	$f(x,z) = P(X=x, Z=z) = \sum_{y=0}^{\infty} f(x,y,z)$
marginal	(Y,Z)	$f(y,z) = P(Y=y, Z=z) = \sum_{x=0}^{\infty} f(x,y,z)$
marginal	X	$f(x) = P(X=x) = \sum_{y=0}^{\infty}\sum_{z=0}^{\infty} f(x,y,z)$
marginal	Y	$f(y) = P(Y=y) = \sum_{x=0}^{\infty}\sum_{z=0}^{\infty} f(x,y,z)$
marginal	Z	$f(z) = P(Z=z) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x,y,z)$
condicional	$(X,Y) \mid Z$	$f(x,y \mid z) = P(X=x, Y=y \mid Z=z) = f(x,y,z)/f(z)$
condicional	$(X,Z) \mid Y$	$f(x,z \mid y) = P(X=x, Z=z \mid Y=y) = f(x,y,z)/f(y)$
condicional	$(Y,Z) \mid X$	$f(y,z \mid x) = P(Y=y, Z=z \mid X=x) = f(x,y,z)/f(x)$
condicional	$X \mid (Y,Z)$	$f(x \mid y,z) = P(X=x \mid Y=y, Z=z) = f(x,y,z)/f(y,z)$
condicional	$Y \mid (X,Z)$	$f(y \mid x,z) = P(Y=y \mid X=x, Z=z) = f(x,y,z)/f(x,z)$
condicional	$Z \mid (X,Y)$	$f(z \mid x,y) = P(Z=z \mid X=x, Y=y) = f(x,y,z)/f(x,y)$
condicional	$X \mid Y$	$f(x \mid y) = P(X=x \mid Y=y) = f(x,y)/f(y)$
condicional	$X \mid Z$	$f(x \mid z) = P(X=x \mid Z=z) = f(x,z)/f(z)$
condicional	$Y \mid X$	$f(y \mid x) = P(Y=y \mid X=x) = f(x,y)/f(x)$
condicional	$Y \mid Z$	$f(y \mid z) = P(Y=y \mid Z=z) = f(y,z)/f(z)$
condicional	$Z \mid X$	$f(z \mid x) = P(Z=z \mid X=x) = f(x,z)/f(x)$
condicional	$Z \mid Y$	$f(z \mid y) = P(Z=z \mid Y=y) = f(y,z)/f(y)$

Tabla 2.7. Diversas funciones de densidad de un vector aleatorio discreto tridimensional.

Distribución multinomial. La distribución multinomial es una generalización de la binomial, de modo que en cada ensayo hay $r \geq 2$ resultados posibles. En el caso de la distribución binomial ($r = 2$), intervienen dos variables aleatorias, que finalmente se reduce a una: el número de éxitos y el número de fracasos en n ensayos. Si $X_1 \sim \text{bin}(n, p)$, entonces X_1 se identifica como el número de éxitos en n ensayos de Bernoulli mientras $X_2 = n - X_1$ representa el total de fracasos, con $X_2 \sim \text{bin}(n, 1-p)$ y $X_1 + X_2 = n$.

Para el caso general de $r \geq 2$, en cada uno de los n ensayos independientes, el resultado se clasifica en uno de los eventos de la partición del espacio muestral Ω: $E_1, \ldots, E_r$. Por lo que estos eventos son excluyentes y su unión es Ω:

$$E_i \cap E_j = \varnothing; \quad \text{para} \quad i \neq j \quad \text{y} \quad E_1 \cup \cdots \cup E_r = \Omega.$$

Para un experimento de $n \geq 1$ ensayos, defina

$$X_i = \text{número de veces que ocurrió el evento } E_i; \quad i = 1, \ldots, r.$$

Así

$$0 \leq X_1, \ldots, X_r \leq n \quad \text{y} \quad X_1 + \cdots + X_r = n.$$

La probabilidad del evento E_i es $p_i = P(E_i) > 0$; para $i = 1, \ldots, r$, con

$$p_1 + \cdots + p_r = 1.$$

Sean los enteros $x_1, \ldots, x_n \geq 0$, tal que $x_1 + \cdots + x_r = n$. El evento $\{X_1 = x_1, \ldots, X_r = x_r\}$, significa que los resultados $E_1, \ldots, E_r$ ocurren $x_1, \ldots, x_r$ veces; respectivamente. Así, la función de densidad conjunta de $(X_1, \ldots, X_n)$ es

$$f(x_1, \ldots, x_r) = P(X_1 = x_1, \ldots, X_r = x_r) = \binom{n}{x_1, \ldots, x_r} p_1^{x_1} \cdots p_r^{x_r},$$

donde el *coeficiente multinomial* es

$$\binom{n}{x_1, \ldots, x_r} = \frac{n!}{x_1! \cdots x_r!}.$$

En tal caso, se dice que $(X_1, \ldots, X_r)$ tiene una distribución conjunta *multinomial* de parámetros n, r y $p_1, \ldots, p_r$. Como en el caso binomial. La distribución multinomial involucra formalmente un vector de dimensión $r - 1$. Por ejemplo, la variable aleatoria X_r se determina por $X_1, \ldots, X_{r-1}$. Por último, la distribución marginal de cada variable aleatoria X_i es binomial:

$$X_i \sim \text{bin}(n, p = p_i); \quad i = 1, \ldots, r.$$

Para el ejemplo de lanzar un dado y observar su cara superior, se tienen $r = 6$ y

$$E_i: \text{ el número de la cara superior es } i,$$
$$p_i = P(E_i) = \frac{1}{6}; \quad i = 1, \ldots, 6.$$

Así

$$
\begin{aligned}
X_i &= \text{número de veces que ocurre } E_i \\
&\sim \text{bin}(n, p = 1/6), \\
n &= X_1 + X_2 + X_3 + X_4 + X_5 + X_6.
\end{aligned}
$$

La función de densidad conjunta es

$$
\begin{aligned}
f(x_1, \ldots, x_6) = P(X_1 = x_1, \ldots, X_6 = x_6) &= \binom{n}{x_1, \ldots, x_6} p_1^{x_1} \cdots p_6^{x_6} \\
&= \binom{n}{x_1, \ldots, x_6} \left(\frac{1}{6}\right)^{x_1} \cdots \left(\frac{1}{6}\right)^{x_6} = \binom{n}{x_1, \ldots, x_6} \frac{1}{6^{x_1 + \cdots + x_6}} \\
&= \binom{n}{x_1, \ldots, x_6} \frac{1}{6^n}; \quad \text{para} \quad 0 \leq x_1, \ldots, x_6 \leq n, \quad \text{con} \quad x_1 + \cdots + x_6 = n.
\end{aligned}
$$

Por argumentos de simetría, las 6 variables aleatorias tienen la misma distribución marginal:

$$
X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} X_3 \stackrel{d}{=} X_4 \stackrel{d}{=} X_5 \stackrel{d}{=} X_6 \sim \text{bin}\left(n, p = \frac{1}{6}\right).
$$

Para el caso general, se deducirá es la distribución marginal de la variable aleatoria X_1; pues la distribución marginal de las demás variables aleatorias se obtiene de manera similar. El conjunto de valores posibles X_1 es

$$
X_1 = 0, \ldots, n.
$$

Como $E_1^c = E_2 \cup \cdots \cup E_r$, entonces

$$
\begin{aligned}
X_1 &: \text{ número de ensayos donde ocurre } E_1, \\
n - X_1 &: \text{ número de ensayos donde ocurre } E_1^c, \\
&= X_2 + \cdots + X_r
\end{aligned}
$$

Por lo que

$$
X_1 \sim \text{bin}(n, p_1) \quad \text{y} \quad X_2 + \cdots + X_r = n - X_1 \sim \text{bin}(n, p_2 + \cdots + p_r).
$$

Este hecho se verifica también analíticamente: para $0 \leq x_1 \leq n$, se tiene

$$\begin{aligned}
f(x_1) &= \sum_{\substack{0\leq x_2,\ldots,x_r\leq n-x_1\\ x_2+\cdots+x_r=n-x_1}} f(x_1,\ldots,x_r)\\
&= \sum_{\substack{0\leq x_2,\ldots,x_r\leq n-x_1\\ x_2+\cdots+x_r=n-x_1}} \binom{n}{x_1,\ldots,x_r} p_1^{x_1}p_2^{x_2}\cdots p_r^{x_r}\\
&= \sum_{\substack{0\leq x_2,\ldots,x_r\leq n-x_1\\ x_2+\cdots+x_r=n-x_1}} \frac{n!}{x_1!x_2!\cdots x_r!} p_1^{x_1}p_2^{x_2}\cdots p_r^{x_r}\\
&= \frac{n!}{x_1!(n-x_1)!}p_1^{x_1} \sum_{\substack{0\leq x_2,\ldots,x_r\leq n-x_1\\ x_2+\cdots+x_r=n-x_1}} \frac{(n-x_1)!}{x_2!\cdots x_r!} p_2^{x_2}\cdots p_r^{x_r}\\
&= \binom{n}{x_1} p_1^{x_1}(1-p_1)^{x_2+\cdots+x_r}\\
&\sum_{\substack{0\leq x_2,\ldots,x_r\leq n-x_1\\ x_2+\cdots+x_r=n-x_1}} \binom{n-x_1}{x_2,\ldots,x_r}\left(\frac{p_2}{1-p_1}\right)^{x_2}\cdots\left(\frac{p_r}{1-p_1}\right)^{x_r}\\
&= \binom{n}{x_1} p_1^{x_1}(1-p_1)^{n-x_1}; \quad \text{para} \quad x_1 = 1,\ldots,n.
\end{aligned}$$

La última suma vale uno, pues sus términos corresponden a la función de densidad de una distribución multinomial de parámetros $n-x_1$, $r-1$ y probabilidades $p_2/(1-p_1),\ldots,p_r/(1-p_1)$. Aquí

$$\frac{p_2}{1-p_1} + \cdots + \frac{p_r}{1-p_1} = \frac{p_2+\cdots+p_r}{1-p_1} = \frac{1-p_1}{1-p_1} = 1.$$

2.9. Aplicación en finanzas: modelo binomial de precios*

En esta sección se analizará el modelo binomial de precios de activos con riesgo.

El modelo binomial de precios es un *proceso estocástico* discreto a tiempo discreto $\{S_0, S_1, \ldots\}$ de un *activo con riesgo*, ya sea la acción de una empresa, precio del barril de petróleo, precio del euro. El proceso es observado con cierta periodicidad; por hora, día, semana, mes. Por ejemplo

$$S_n = \text{precio del barril de crudo al día } n; \quad n = 0, 1, \ldots$$

El modelo binomial de precios asume que entre dos día consecutivos el precio del barril sube u unidades o baja d veces, con $S_0 > 0$ fijo y

$$0 < d < 1 < u.$$

Esto es

$$
\begin{aligned}
S_1 &= \begin{cases} S_0 u & \text{sube} \\ S_0 d & \text{baja,} \end{cases} \\
S_2 &= \begin{cases} S_1 u & \text{sube} \\ S_1 d & \text{baja,} \end{cases} \\
&= \begin{cases} S_0 u^2 & \text{sube y sube} \\ S_0 ud & \text{sube y baja ó baja y sube} \\ S_0 d^2 & \text{baja y baja,} \end{cases} \\
&\vdots \\
S_{n+1} &= \begin{cases} S_n u & \text{si sube último precio} \\ S_n d & \text{si baja último precio} \end{cases} \\
&\vdots
\end{aligned}
$$

Considere *éxito* cuando suba el precio del activo con riesgo, con probabilidad

$$p = P(E) = P(\text{precio sube}).$$

Al asumir independencia de los cambios de los precios en cada día, se tienen n ensayos de Bernoulli. De esta manera, al definir la variable aleatoria

$$X_n : \text{número de alzas del precio del barril entre los tiempos } 0 \text{ y } n,$$

entonces X_n tienen distribución binomial de parámetros n y p. Al tiempo n, los valores posibles de S_n son

$$S_n = S_0 d^n, S_0 u d^{n-1}, \ldots, S_0 u^{n-1} d, S_0 u^n,$$

los cuales corresponden a los propios de X_n:

$$X_n = 0, 1, \ldots, n-1, n.$$

Note que

$$S_n = S_0 u^{X_n} d^{n-X_n} = S_0 d^n \left(\frac{u}{d}\right)^{X_n}. \tag{2.97}$$

Como X_n es una variable aleatoria discreta, entonces S_n también es una variable aleatoria discreta. Su función de densidad es

$$P\left[S_n = S_0 d^n \left(\frac{u}{d}\right)^x\right] = P(X_n = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, \ldots, n.$$

En particular, si $d = 1/u$, entonces la expresión (2.97) se simplifica como

$$S_n = S_0 d^n \left(\frac{u}{d}\right)^{X_n} = S_0 \left(\frac{1}{u}\right)^n \left(\frac{u}{1/u}\right)^{X_n} = S_0 u^{2X_n - n}.$$

Por ejemplo, suponga

$$S_0 = 200, \quad n = 10, \quad u = 1.01 \quad \text{y} \quad d = \frac{1}{u} \approx 0.9901. \tag{2.98}$$

Las ganancias y pérdidas son del orden de 1 %: $u - 1 = 0.01 = 1\,\%$ y $d - 1 = (1-u)/u = -0.01/1.01 \approx -0.0099 = -0.99\,\%$. Así

$$S_1 = \begin{cases} 202 \\ 198.0198, \end{cases} \qquad S_2 = \begin{cases} 204.02 \\ 200 \\ 196.06, \end{cases} \qquad S_3 = \begin{cases} 206.0602 \\ 202 \\ 198.0198 \\ 194.118, \end{cases} \qquad \ldots$$

El espacio muestral es

$$\Omega = \text{conjunto de trayectorias} = \{\omega = (\omega_1, \ldots, \omega_{10}) : \omega_i = 0, 1;\ i = 1, \ldots, 10\}.$$

La Figura 2.16[izquierda] muestra los nodos de las $2^n = 2^{10} = 1024$ diferentes trayectorias del espacio muestral, para $n = 10$. Cada trayectoria representa un evento simple. En contraste, si sólo se registra el precio del activo con riesgo al tiempo $n = 10$, entonces S_n tiene $n+1 = 11$ diferentes resultados posibles, como se muestra en la parte [derecha] de la misma figura.

Al retomar el modelo general (2.97), la esperanza y varianza de S_n se obtienen vía la FGP de X_n. Por (2.87) recuerde

$$\mathrm{E}\, t^{X_n} = [pt + 1 - p]^n; \quad t \geq 0.$$

Entonces

$$\begin{aligned}
\mathrm{E}\, S_n &= \mathrm{E}\left[S_0 d^n \left(\frac{u}{d}\right)^{X_n}\right] = S_0 d^n\, \mathrm{E}\left(\frac{u}{d}\right)^{X_n} = S_0 d^n \left[p\left(\frac{u}{d}\right) + 1 - p\right]^n, \\
\mathrm{E}\, S_n^2 &= \mathrm{E}\left[S_0 d^n \left(\frac{u}{d}\right)^{X_n}\right]^2 = S_0^2 d^{2n}\, \mathrm{E}\left(\frac{u}{d}\right)^{2X_n} = S_0^2 d^{2n} \left[p\left(\frac{u}{d}\right)^2 + 1 - p\right]^n, \\
\mathrm{V}\, S_n &= \mathrm{E}\, S_n^2 - [\mathrm{E}\, S_n]^2 = S_0^2 d^{2n} \left(\left[p\left(\frac{u}{d}\right)^2 + 1 - p\right]^n - \left[p\left(\frac{u}{d}\right) + 1 - p\right]^{2n}\right), \\
\mathrm{d.\,e.}(S_n) &= \sqrt{\mathrm{V}\, S_n} = S_0 d^n \left(\left[p\left(\frac{u}{d}\right)^2 + 1 - p\right]^n - \left[p\left(\frac{u}{d}\right) + 1 - p\right]^{2n}\right)^{1/2}.
\end{aligned}$$

En particular, para $d = 1/u$, se tiene

$$\begin{aligned}
\mathrm{E}\, S_n &= S_0 u^{-n}(pu^2 + 1 - p)^n = S_0 \left(pu + \frac{1-p}{u} \right)^n, \\
\mathrm{E}\, S_n^2 &= S_0^2 u^{-2n}(pu^4 + 1 - p)^n = S_0^2 \left(pu^2 + \frac{1-p}{u^2} \right)^n, \\
\mathrm{V}\, S_n &= S_0^2 u^{-2n} \left[(pu^4 + 1 - p)^n - (pu^2 + 1 - p)^{2n} \right] \\
&= S_0^2 \left[\left(pu^2 + \frac{1-p}{u^2} \right)^n - \left(pu + \frac{1-p}{u} \right)^{2n} \right], \\
\mathrm{d.\,e.}(S_n) &= S_0 u^{-n} \left[(pu^4 + 1 - p)^n - (pu^2 + 1 - p)^{2n} \right]^{1/2} \\
&= S_0 \left[\left(pu^2 + \frac{1-p}{u^2} \right)^n - \left(pu + \frac{1-p}{u} \right)^{2n} \right]^{1/2}.
\end{aligned}$$

Por ejemplo, considere (2.98) y $p = 0.5$. Entonces

$$\begin{aligned}
\mathrm{E}\, S_n &= S_0 \left(pu + \frac{1-p}{u} \right)^n = 200 \left(0.5(1.01) + \frac{0.5}{1.01} \right)^n = 200 \times 1.00005^n, \\
\mathrm{d.\,e.}(S_n) &= S_0 \left[\left(pu^2 + \frac{1-p}{u^2} \right)^n - \left(pu + \frac{1-p}{u} \right)^{2n} \right]^{1/2} \\
&= 200 \left[\left(0.5(1.01^2) + \frac{0.5}{1.01^2} \right)^n - 1.00005^{2n} \right]^{1/2} \\
&= 200 \left[1.000198^n - 1.000099^n \right]^{1/2}.
\end{aligned}$$

En la Figura 2.17 se muestra la gráfica de precio promedio de activo con riesgo $\mathrm{E}\, S_n$, así como el intervalo $\mathrm{E}\, S_n \pm 1.96 \times \mathrm{d.\,e.}(S_n)$; para $n = 0, \ldots, 30$, bajo el modelo binomial, de este ejemplo. Para un tiempo fijo $n \geq 0$, el precio S_n se encuentra dentro del intervalo $\mathrm{E}\, S_n \pm 1.96 \times \mathrm{d.\,e.}(S_n)$, con una probabilidad aproximada de 0.95.

2.10. Sumas aleatorias*

Considere la *suma aleatoria*

$$X = \xi_1 + \cdots + \xi_N,$$

donde $\xi_1, \xi_2, \xi_3, \ldots,$ es una colección de variables aleatorias independientes e idénticamente distribuidas y N es una variable aleatoria discreta, con función de densidad

$$f(n) = P(N = n); \quad \text{para} \quad n \geq 0.$$

Si $N = 0$, la suma es cero. Este modelo tiene muchas aplicaciones, que a continuación se muestran algunas.

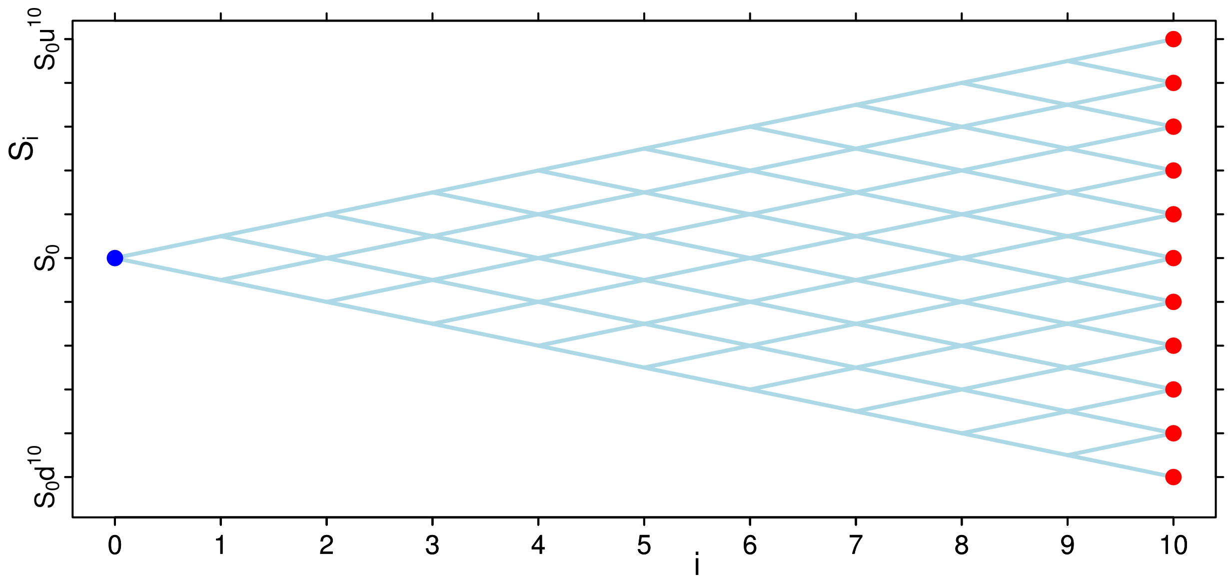


Figura 2.16. Espacio muestral Ω del modelo binomial de precios, con $\#\Omega = 2^n = 2^{10} = 1024$ trayectorias o eventos simples. Los puntos rojos son $n + 1 = 11$ eventos simples del espacio muestral alterno.

1. Teoría de colas (queueing theory). Sea N el número de clientes que llegan a un servidor (banco, tortillería, área de pintado de una planta automotriz), durante un día. Sea ξ_i el tiempo requerido para atender al i-ésimo cliente. Entonces, $X = \xi_1 + \cdots + \xi_N$ es el tiempo total de servicio.

2. Riesgo de seguros (insurance). Suponga que en una semana, cierta compañía aseguradora recibe N reclamaciones, con ξ_i el monto de la i-ésima reclamación. Entonces la suma de reclamaciones es $X = \xi_1 + \cdots + \xi_N$.

3. Modelo de población. Considere

$$\begin{aligned} N &= \text{número de insectos hembras de una especie de cierta generación,} \\ \xi_i &= \text{número de descendientes del } i\text{-ésimo insecto,} \\ X &= \text{total de descendientes en una generación.} \end{aligned}$$

Suponga que

$$\mathrm{E}\,\xi = \mu \quad \text{y} \quad \mathrm{V}\,\xi = \sigma^2.$$

Entonces

$$\mathrm{E}\,X = \mu \cdot \mathrm{E}\,N \tag{2.99}$$

y

$$\mathrm{V}\,X = \mu^2\,\mathrm{V}\,N + \sigma^2\,\mathrm{E}\,N. \tag{2.100}$$

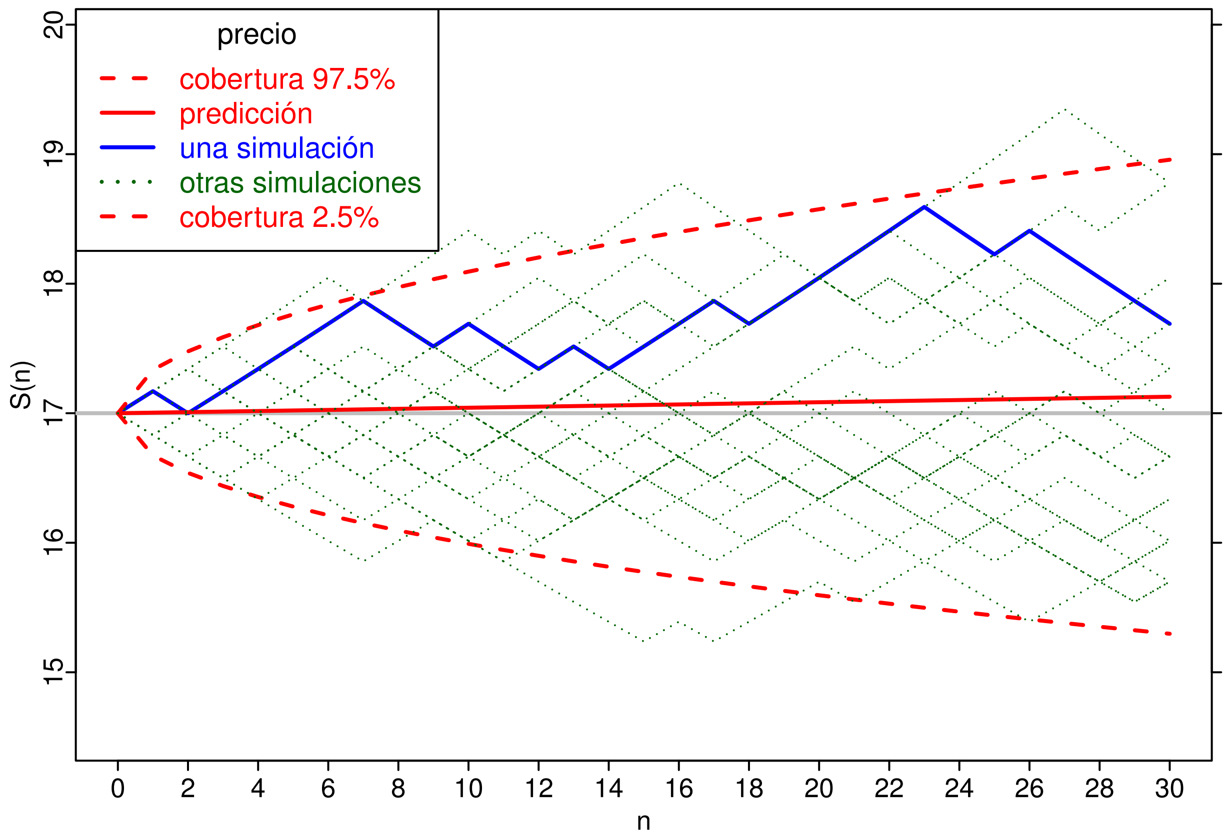


Figura 2.17. Gráfica de precio promedio de activo con riesgo $\mathrm{E}\, S_n$, así como banda $\mathrm{E}\, S_n \pm 1.96 \times \mathrm{d.\,e.}(S_n)$; $n = 0, 1, \ldots, 30$, bajo el modelo binomial.

La expresión (2.99) se conoce como identidad de *Wald*, que se deduce de la ley de esperanza iterada (2.94), según

$$\begin{aligned}
\mathrm{E}\,X &= \mathrm{E}\{\mathrm{E}[X \mid N]\} = \sum_{n=0}^{\infty} \mathrm{E}[X \mid N = n] f(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[\xi_1 + \cdots + \xi_N \mid N = n] f(n) = \sum_{n=0}^{\infty} \mathrm{E}[\xi_1 + \cdots + \xi_n \mid N = n] f(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[\xi_1 + \cdots + \xi_n] f(n) = \sum_{n=0}^{\infty} n\,\mathrm{E}(\xi_1) f(n) \\
&= \mu \sum_{n=0}^{\infty} n f(n) = \mu\,\mathrm{E}\,N.
\end{aligned}$$

Para la varianza, observe que

$$\begin{aligned}
\mathrm{V}\,X &= \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}(X - \mu\,\mathrm{E}\,N)^2 \\
&= \mathrm{E}(X - \mu N + \mu N - \mu\,\mathrm{E}\,N)^2 \\
&= \mathrm{E}(X - \mu N)^2 + \mathrm{E}(\mu N - \mu\,\mathrm{E}\,N)^2 + 2\,\mathrm{E}[(X - \mu N)(\mu N - \mu\,\mathrm{E}\,N)] \\
&= \mathrm{E}(X - \mu N)^2 + \mu^2\,\mathrm{E}(N - \mathrm{E}\,N)^2 + 2\mu\,\mathrm{E}[(X - \mu N)(N - \mathrm{E}\,N)] \\
&= \mathrm{E}(X - \mu N)^2 + \mu^2 \cdot \mathrm{V}\,N + 2\mu\,\mathrm{E}[(X - \mu N)(N - \mathrm{E}\,N)].
\end{aligned}$$

Al aplicar la ley de esperanza iterada (2.94) a $(X - N\mu)^2$, se obtiene

$$\mathrm{E}(X - N\mu)^2 = \mathrm{E}(\mathrm{E}[(X - N\mu)^2 \mid N])$$

$$\begin{aligned}
\mathrm{E}(X - N\mu)^2 &= \mathrm{E}\{\mathrm{E}[(X - N\mu)^2 \mid N]\} \\
&= \sum_{n=0}^{\infty} \mathrm{E}[(X - N\mu)^2 \mid N = n] f(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[\{\xi_1 + \cdots + \xi_n - n\mu\}^2 \mid N = n] f(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[\{(\xi_1 - \mu) + \cdots + (\xi_n - \mu)\}^2] f(n) \\
&= \sum_{n=0}^{\infty} \mathrm{E}[(\xi_1 - \mu)^2 + \cdots + (\xi_n - \mu)^2] f(n) \\
&= \sum_{n=0}^{\infty} n\sigma^2 \cdot f(n) = \sigma^2 \sum_{n=0}^{\infty} n f(n) \\
&= \sigma^2\,\mathrm{E}\,N.
\end{aligned}$$

De manera similar, se comprueba (2.100):

$$\begin{aligned}
\mathrm{E}[(X - \mu N)(N - \mathrm{E}\,N)] &= \mathrm{E}\{\mathrm{E}[(X - \mu N)(N - \mathrm{E}\,N) \mid N]\} \\
&= \mathrm{E}\{(N - \mathrm{E}\,N)\,\mathrm{E}[X - \mu N \mid N]\} \\
&= \mathrm{E}\{(N - \mathrm{E}\,N)(\mu N - \mu N)\} = 0.
\end{aligned}$$

La variable aleatoria X es discreta o continua, siempre que ξ lo sea. Para conocer la distribución de X, se verán algunos resultados generales.

Sea (X, N) un vector aleatorio tal que N es una variable aleatoria discreta y X continua. La función de distribución condicional de X dado $N = n$ es

$$F(x \mid N = n) = P(X \leq x \mid N = n) = \frac{P(X \leq x, N = n)}{P(N = n)}.$$

Note que

$$F(x \mid N = n) = \int_{-\infty}^{x} f(u \mid N = n) \, \mathrm{d}\, u,$$

donde la función de densidad condicional de X dado $N = n$ es

$$f(x \mid N = n) = \frac{\mathrm{d}}{\mathrm{d}x} F(x \mid N = n).$$

La respectiva esperanza condicional de X dado $N = n$ es

$$\mathrm{E}[X \mid N = n] = \int_{-\infty}^{\infty} x f(x \mid N = n) \, \mathrm{d}\, x.$$

La función de densidad conjunta se representa como

$$f(x, n) = f(x \mid N = n) f(n),$$

donde el conjunto de vectores posibles (x, n) es tal que

$$-\infty < x < \infty, \quad n = 0, 1, \ldots$$

Por la ley de probabilidad total (1.19), la distribución marginal de la variable aleatoria X satisface

$$\begin{aligned}
F(x) = P(X \leq x) &= \sum_{n=0}^{\infty} P(X \leq x, N = n) \\
&= \sum_{n=0}^{\infty} \left[\int_{-\infty}^{x} f(u \mid N = n) \, \mathrm{d}\, u \right] f(n) = \int_{-\infty}^{x} \sum_{n=0}^{\infty} f(u \mid N = n) f(n) \, \mathrm{d}\, u \\
&= \int_{-\infty}^{x} \sum_{n=0}^{\infty} f(u, n) \, \mathrm{d}\, u.
\end{aligned}$$

La respectiva función de densidad marginal es

$$f(x) = F'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \int_{-\infty}^{x} \sum_{n=0}^{\infty} f(u, n) \, \mathrm{d}\, u = \sum_{n=0}^{\infty} f(x, n).$$

2.11. Ejercicios

Esta es la sección de ejercicios del capítulo de variables aleatorias discretas.

1. Luego de ir a un baño público, el 37.04 % de las personas adultas no se lavan las manos con jabón o gel antibacterial. Sea la variable aleatoria X, el total de personas que no se lavan bien las manos, de un grupo de 7 personas que sale del baño. Obtenga:

   a) El modelo de distribución de esta variable aleatoria

   b) El conjunto de valores posibles de la variable aleatoria X.

   c) La función de densidad $f(x) = P(X = x)$; para $x = 0, 1, 2, \dots$

   d) ¿Cuáles son los dos valores posibles más factibles de X, y los dos menos factible?

   e) ¿Cuál es la probabilidad de que al menos una de esas 7 personas no se haya lavado las manos?

   f) El valor esperado $\mu = \mathrm{E}\, X$

   g) La varianza $\sigma^2 = \mathrm{V}\, X = \mathrm{E}(X - \mu)^2$

   h) La desviación estándar $\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{V}\, X}$

   i) La gráfica de la función de densidad $f(x)$. En el eje horizontal de la gráfica identifique el valor esperado μ y la moda (2.38)

   j) La gráfica de la función de distribución $F(x) = P(X \leq x)$; para $x \in \mathbf{R}$. En el eje horizontal de la gráfica identifique el valor esperado μ y la mediana (2.37).

2. Continuación del Ejemplo 2.1.1. Considere la variable aleatoria X, el número de juegos ganados al apostar 5 veces a los números rojos en la ruleta. Sea la variable aleatoria Y la ganancia obtenida al apostar un peso en cada juego. Obtenga

   a) El conjunto de valores posibles de la variable aleatoria Y

   b) La relación funcional de Y en términos de X: $Y = g(X)$

   c) La ganancia esperada: $\mathrm{E}\, Y = \mathrm{E}\, g(X)$.

3. Considere un jugador de la ruleta con estrategia de apostar un número impar de n veces a un evento de probabilidad $0 < p < 1/2$. Para fijar el evento de interés, apóyese en el paño de la ruleta de la Figura 1.4. Sean las variables aleatorias X el número de juegos ganados y Y la respectiva ganancia. Obtenga

   a) La función de densidad de ambas variables aleatorias, así como su gráfica.

   b) La función de distribución de ambas variables aleatorias, así como su gráfica.

   c) La probabilidad de ganar $\{X > n/2\}$.

   d) El número esperado de éxitos $\mathrm{E}\, X$, junto con la ganancia esperada $\mathrm{E}\, Y$.

   e) Obtenga la gráfica respecto de n de los indicadores: $P(X > n/2)$, $\mathrm{E}\, X$ y $\mathrm{E}\, Y$.

4. Sea X una variable aleatoria con segundo momento finito (no necesariamente discreta). Considere la relación $a \to \mathrm{E}(X-a)^2$; para $a \in \mathbf{R}$. Compruebe que el mínimo de esta función es $a = \mathrm{E}\, X = \mu$. Este punto corresponde a la varianza de la variable aleatoria X.

5. Sea Y una variable aleatoria con distribución geométrica:
$$P(Y=y) = p(1-p)^y; \quad \text{para} \quad y = 0, 1, 2, \ldots$$
Compruebe que su función de distribución es
$$F(y) = P(Y \leq y) = 1 - (1-p)^{y+1}; \quad \text{para} \quad y = 0, 1, 2, \ldots$$
Como consecuencia, la función de *supervivencia* resulta
$$P(Y > y) = (1-p)^{y+1}; \quad \text{para} \quad y = 0, 1, 2, \ldots$$

6. Sea (X, Y) un vector aleatorio discreto, con conjunto de valores posibles $x, y = 0, 1, \ldots$ Compruebe que X y Y son variables aleatorias independientes, si y sólo si la función de densidad conjunta es el producto de las densidades marginales (2.47).

7. Compruebe que las desigualdades de momentos 4., $4b$. y $4c$., del Teorema 2.4.4, se deducen por $4a$.

8. Sea X una variable aleatoria de distribución uniforme discreta en los enteros consecutivos $\{a, a+1, \ldots, b\}$, con $a \leq b$. Obtenga su función de densidad $f(x) = P(X = x)$; para $x = a, a+1, \ldots, b$. Compruebe que la media y varianza de X son, respectivamente
$$\mathrm{E}\, X = \frac{a+b}{2} \quad \text{y} \quad \mathrm{V}\, X = \frac{(b-a+1)^2 - 1}{12}.$$
Sugerencia: desarrolle primero el caso $a = 1$.

9. Sea X una variable aleatoria de distribución binomial de parámetros n y p. Demuestre que $Y \sim \mathrm{bin}(n, 1-p)$. Compruebe además que $\rho(X, Y) = -1$.

10. Verifique que la correlación de X y Y coincide con la correlación y covarianza de sus correspondientes variables estandarizadas:
$$\rho(X, Y) = \rho\left(\frac{X-\mu_1}{\sigma_1}, \frac{Y-\mu_2}{\sigma_2}\right) = \mathrm{C}\left(\frac{X-\mu_1}{\sigma_1}, \frac{Y-\mu_2}{\sigma_2}\right),$$
donde
$$\mu_1 = \mathrm{E}\, X, \quad \mu_2 = \mathrm{E}\, Y, \quad \sigma_1^2 = \mathrm{V}\, X, \quad \sigma_2^2 = \mathrm{V}\, Y.$$

11. Considere el vector aleatorio (X_1, X_2, X_3), de distribución trinomial (multinomial con $r = 3$), de parámetros $n \geq 1$ y $p_1, p_2, p_3 \geq 0$, con
$$p_1 + p_2 + p_3 = 1.$$
Compruebe que
$$X_1 \sim \mathrm{bin}(n, p_1), \quad X_2 \sim \mathrm{bin}(n, p_2), \quad X_3 \sim \mathrm{bin}(n, p_3).$$

12. Encontrar la función de densidad de una variable aleatoria discreta que no tenga esperanza. Sugerencia: considere la convergencia y divergencia de las siguientes dos series

$$\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6} \quad \text{y} \quad \sum_{x=1}^{\infty} \frac{1}{x} = \infty.$$

13. Sea X una variable aleatoria. Compruebe la equivalencia

$$\mathrm{E}\,|X-\mu|^r < \infty \quad \Leftrightarrow \quad \mathrm{E}\,|X|^r < \infty; \quad r \geq 1.$$

14. Demuestre que la desigualdad de Cauchy Bunyakovsky Schwarz (2.62), implica la de Rogers Hölder (2.60). Sugerencia: véase demostración de $\underline{(C_1) \Rightarrow (H_0)}$ en [16, Teorema 2.2, p.p. 7].

15. Con la función generadora de probabilidad, obtenga el sesgo y la curtosis de la distribución geométrica. Sugerencia: aplique expresiones (2.81)-(2.84).

16. Sea X una distribución binomial de parámetros $n \geq 1$ y $0 \leq p \leq 1$. Con la técnica de la función generadora de momentos, verifique que el coeficiente de curtosis es (2.88).

17. Sea X una variable aleatoria de distribución binomial negativa de parámetros $r \geq 1$ y $0 < p < 1$. Demuestre que las funciones generadoras de momentos y de probabilidad son, respectivamente

$$M(t) = \mathrm{E}\,e^{tX} = \left(\frac{pe^t}{1-(1-p)e^t}\right)^r; \quad \text{para} \quad t < -\log(1-p)$$

y

$$\phi(t) = \mathrm{E}\,t^X = \left(\frac{pt}{1-(1-p)t}\right)^r; \quad \text{para} \quad t < \frac{1}{1-p}.$$

18. Sea X una variable aleatoria de distribución uniforme discreta en $x_1, x_2, x_3 \in \mathbf{R}$, con $x_1 < x_2 < x_3$. Compruebe que curtosis$(X) = 1.5$. Sugerencia: sin pérdida de generalidad, asuma el caso estándar; $\mathrm{E}\,X = 0$ y $\mathrm{V}\,X = 1$.

19. Sean X y Y dos variables aleatorias independientes. Compruebe que también son independientes las variables aleatorias e^X y e^Y.

20. Sea X una variable aleatoria con función generadora de momentos $M(t) = \mathrm{E}\,e^{tX}$, definida en cierto intervalo abierto que contiene a cero. Demuestre

   - $|X|$ tiene función generadora de momentos definida en dicho intervalo. Sugerencia: considere la desigualdad

   $$e^{t|x|} \leq e^{tx} + e^{-tx}; \quad \text{para todo} \quad x, t \in \mathbf{R}.$$

   ¿Qué puede decir del recíproco?

- $M(t)$ es función convexa en su dominio: para todo $0 \leq \alpha \leq 1$ y $s, t \in \mathbf{R}$, resulta

$$M(\alpha s + (1-\alpha)t) \leq \alpha M(s) + (1-\alpha)M(t).$$

¿Qué puede decir del signo de la segunda derivada de $M(t)$?

- $M(t)$ es función log convexa en su dominio, es decir, es convexa la función $\log M(t)$. Sugerencia: compruebe

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\{\log M(t)\} = \frac{M''(t)}{M(t)} - \left(\frac{M'(t)}{M(t)}\right)^2 = \mathrm{E}\left[X^2 \frac{e^{tX}}{M(t)}\right] - \left(\mathrm{E}\left[X \frac{e^{tX}}{M(t)}\right]\right)^2.$$

Luego, la ultima expresión es la varianza de cierta variable aleatoria con función de densidad $g(x) = e^{tx} f(x)/M(t)$.

21. Probe la igualdad

$$\mathrm{V}[Y \mid X = x] = \mathrm{E}[Y^2 \mid X = x] - (\mathrm{E}[Y \mid X = x])^2; \quad \text{para} \quad x \in \mathbf{R}.$$

Esta propiedad se escribe de manera equivalente como

$$\mathrm{V}[Y \mid X] = \mathrm{E}[Y^2 \mid X] - (\mathrm{E}[Y \mid X])^2.$$

22. Sea (X, Y) un vector aleatorio con $\mathrm{V}\, Y < \infty$. Pruebe la identidad (2.95), que relaciona la varianza de Y con la de $\mathrm{E}[Y \mid X]$:

$$\mathrm{V}\, Y = \mathrm{E}\{\mathrm{V}[Y \mid X]\} + \mathrm{V}\{\mathrm{E}[Y \mid X]\}.$$

23. Desarrolle el modelo binomial para las ganancias en el juego de la ruleta. Recuerde que la probabilidad de éxito es $P(\text{rojo}) = p = 18/38$.

Capítulo 3

Variables aleatorias continuas

En este capítulo se describe a las variables aleatorias continuas junto con sus propiedades básicas. Se abordarán los modelos de distribución de probabilidad normal, exponencial, uniforme continua, gamma y Weibull.

3.1. Introducción a las variables aleatorias continuas

Informalmente, una variable aleatoria es continua si toma cualquier valor en un intervalo. Por ejemplo, la temperatura que habrá el día de mañana a la 08:00 horas se encuentra en un rango amplio de valores de la recta real, como $(-30, 50)$. Otro ejemplo de variable aleatoria continua es el error en el ancho de un tornillo con especificación de cinco milímetros, seleccionado al azar de un lote. Su conjunto de valores posibles es el intervalo $(-\infty, \infty)$. Las variables aleatorias tiempo de espera en la fila y tiempo de servicio de un cliente en un servidor tiene como conjunto de valores posibles al intervalo $(0, \infty)$.

Una variable aleatoria continua X tiene propiedades similares a las discretas. Por ejemplo, satisface las propiedades de la esperanza, varianza y covarianza (2.4.4) y (2.5.1). Sin embargo, por definición, la probabilidad de que tome un valor específico es cero. En consecuencia, su función de distribución $F(x)$ es continua mientras que su *función de densidad* $f(x)$ no es probabilidad.

Definición 3.1.1. Sea $X : \Omega \to \mathbf{R}$, una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se dice que X es una *variable aleatoria continua* si

$$P(X = x) = 0; \quad \text{para todo} \quad x \in \mathbf{R}. \tag{3.1}$$

Considere el ejemplo de la posición X respecto del pivote de la picadura de un neumático de radio un metro. Esta es una variable aleatoria que tomar cualquier valor en el intervalo $(0, 2\pi)$. Mas sin embargo cada valor posible x en dicho intervalo tiene probabilidad cero de ocurrir. La picadura del neumático ocurre al azar según la distribución *uniforme continua* en el intervalo $(0, 2\pi)$. Se denota como $X \sim \mathrm{U}(a = 0, b = 2\pi)$. Su función de distribución es

$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{2\pi} & 0 \leq x \leq 2\pi \\ 1 & x \geq 2\pi. \end{cases}$$

La probabilidad de que la picadura se encuentre una región dentro del intervalo $(0, 2\pi)$, es proporcional a su longitud. En consecuencia

$$
\begin{aligned}
P(X \leq \pi) = F(\pi) &= \frac{\pi}{2\pi} = \frac{1}{2} = 0.5,\\
P\left(X \leq \frac{2\pi}{3}\right) = F\left(\frac{2\pi}{3}\right) &= \frac{2\pi/3}{2\pi} = \frac{1}{3} = 0.3333,\\
P\left(X \leq \frac{\pi}{2}\right) = F\left(\frac{\pi}{2}\right) &= \frac{\pi/2}{2\pi} = \frac{1}{4} = 0.25,\\
P\left(\pi < X \leq \frac{3\pi}{2}\right) = F\left(\frac{3\pi}{2}\right) - F(\pi) &= \frac{3\pi/2}{2\pi} - \frac{\pi}{2\pi} = \frac{1}{4} = 0.25.
\end{aligned}
$$

Así mismo

$$
\begin{aligned}
P\left(X > \frac{3\pi}{2}\right) &= 1 - P\left(X \leq \frac{3\pi}{2}\right) = 1 - F\left(\frac{3\pi}{2}\right) = 1 - \frac{3\pi/2}{2\pi} = 1 - \frac{3}{4} = 0.25,\\
P(X \leq 7.5) &= P(X \leq 2\pi) = F(7.5) = F(2\pi) = 1,\\
P(X \leq -1.1) &= P(X \leq 0) = F(-1.1) = F(0) = 0.
\end{aligned}
$$

Por el Teorema 2.1.3 y la Proposición 2.1.4, la función de distribución $F(x)$ es continua por la derecha y tiene límite por la izquierda, con

$$F(x-) = P(X < x) \leq P(X \leq x) = F(x) = F(x+); \quad \text{para} \quad x \in \mathbf{R}.$$

Por otro lado, considere la unión de eventos excluyentes

$$\{X \leq x\} = \{X < x\} \cup \{X = x\}.$$

Así

$$F(x) = P(X \leq x) = P(X < x) + P(X = 0) = F(x-) + P(X = 0); \quad \text{para} \quad x \in \mathbf{R}.$$

Note que $P(X = x) = 0$, si y sólo si $F(x) = F(x-)$. En tal caso, la función $F(x)$ es continua en x. Se concluye la equivalencia de las siguientes afirmaciones.

1. $F(x)$ es continua en x
2. $F(x-) = F(x)$
3. $P(X < x) = P(X \leq x)$
4. $P(X = x) = 0$.

El siguiente teorema resume las caracterizaciones de una variable aleatoria continua.

Teorema 3.1.2. Sea X una variable aleatoria, con función de distribución

$$F(x) = P(X \leq x); \quad \text{para} \quad x \in \mathbf{R}.$$

Entonces, las siguientes afirmaciones son equivalentes.
1. X es una variable aleatoria continua.
2.
$$P(X = x) = 0; \quad \text{para} \quad x \in \mathbf{R}.$$
3. La función de distribución $F(x)$ es continua.
4. La función de distribución $F(x)$ es continua por la izquierda.
5.
$$P(X < x) = P(X \leq x); \quad \text{para} \quad x \in \mathbf{R}.$$

3.2. Variables aleatorias absolutamente continuas

La familia de variables aleatorias absolutamente continuas son un caso particular de las continuas. Esta tiene una extendida aplicación en todas las áreas del conocimiento, en la modelación de los fenómenos aleatorios de naturaleza continua.

Sea X una variable aleatoria con función de distribución $F(x) = P(X \leq x)$; para $x \in \mathbf{R}$. Por su propiedad creciente, esta función satisface las siguientes propiedades; véase [11, Teorema 6.1.1, Teorema de Lebsgue 6.2. y Corolario 6.2.4].

- $F(x)$ es continua, excepto por un conjunto de puntos a lo más numerable.
- $F(x)$ es diferenciable *casi en todas partes*.
- Su derivada es integrable, aunque puede no cumplir el teorema fundamental del cálculo.

El caso particular de una variable aleatoria absolutamente continua es cuando $F(x)$ resulta la integral indefinida de su derivada, según el teorema fundamental del cálculo, como se define a continuación.

Se dice que la variable aleatoria X es *absolutamente continua* si su función de distribución es *absolutamente continua*:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t; \quad x \in \mathbf{R}, \tag{3.2}$$

donde, casi en todas partes

$$f(x) \mathrel{\hat{=}} F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x); \quad x \in \mathbf{R}. \tag{3.3}$$

A esta función se le conoce como la *función de densidad* de la variable aleatoria X o de $F(x)$, la cual es única casi en todas partes. El *conjunto de valores posibles* de X es la región de la recta real donde su función de densidad está definida y es positiva:

$$\mathcal{X} = \{x \in \mathbf{R} : f(x) > 0\}.$$

Así que

$$P(X \in \mathcal{X}) = 1 \quad \text{y} \quad P(X \in \mathcal{X}^c) = 0.$$

Como $F(x)$ es una función continua, entonces la variable aleatoria X es continua.

Usualmente, el conjunto de valores posibles de una variable aleatoria absolutamente continua es un intervalo abierto:

$$\mathcal{X} = \{x \in \mathbf{R} : f(x) > 0\} = (a, b), \quad \text{para ciertos} \quad -\infty \leq a < b \leq \infty.$$

En tal caso, la función de distribución $F(x)$ es estrictamente creciente en dicho intervalo y

$$f(x) = F'(x) > 0; \quad \text{para todo} \quad x \in (a, b).$$

Véase Ejercicio 3.1. De aquí en adelante, salvo que se mencione lo contrario, se identificará a una variable absolutamente continua como continua. Cabe recordar que "absolutamente continuo" es un caso particular de "continuo".

Análogo al caso discreto, la función de densidad $f(x)$ determina la distribución de una variable aleatoria continua X. Si bien $f(x)$ no es una probabilidad, esta función tiene toda la información de probabilidad relevante de X antes del experimento. En este sentido, por (2.5) y (3.2), se cumple el teorema fundamental del cálculo:

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \\ &= \int_a^b F'(t)\mathrm{d}t \\ &= \int_a^b f(t)\mathrm{d}t; \quad \text{con} \quad -\infty \leq a \leq b \leq \infty. \end{aligned} \tag{3.4}$$

Además, con la función de densidad $f(x)$ se calcula la probabilidad de cualquier evento asociado a la variable aleatoria X:

$$P(X \in B) = \int_B f(t)\mathrm{d}t; \quad \text{para} \quad B \in \mathcal{B}(\mathbf{R}). \tag{3.5}$$

Las fórmulas (3.2), (3.4) y (3.5) son equivalentes. Por ejemplo con $a = -\infty$ y $b = x$ en (3.4), resulta (3.2). Así mismo, al aplicar (3.5), en los conjuntos de Borel $B = (-\infty, x]$ y $B = (a, b]$, se obtiene (3.2) y (3.4); respectivamente. Por el Ejercicio 3.2, se demuestra la equivalencia entre (3.4) y (3.2). La deducción de (3.5), bajo (3.4) o (3.2) como hipótesis, va más allá del alcance de este libro. Por otro lado, note que (3.4) implica (3.1):

$$P(X = x) = P(x \leq X \leq x) = \int_x^x f(t)\mathrm{d}t = 0; \quad \text{para} \quad x \in \mathbf{R}.$$

Por otro lado, note que las siguientes probabilidades son iguales:

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(t)\mathrm{d}t; \tag{3.6}$$

para $-\infty \leq a \leq b \leq \infty$. El siguiente teorema caracteriza con dos propiedades a la función de densidad de una variable aleatoria continua.

Teorema 3.2.1. Sea X una variable aleatoria continua, con función de densidad (3.3): $f(x) = F'(x)$; para $x \in \mathbf{R}$. Entonces se satisface

1.
$$f(x) \geq 0; \quad \text{para} \quad x \in \mathbf{R}. \tag{3.7}$$

2. La función $f(x)$ es integrable e
$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1. \tag{3.8}$$

Recíprocamente, si $f(x)$ es una función real que cumpla ambas propiedades, entonces esta corresponde a la función densidad de cierta variable aleatoria continua X. En tal caso, la respectiva función de distribución $F(x)$ satisface (3.3)-(3.5).

Demostración. ($\Rightarrow$). La propiedad (3.7) se deduce del hecho que la función de distribución $F(x)$ es creciente en el sentido amplio y derivable en la recta real. Así, su derivada es no negativa. Por otro lado, (3.8) se obtiene al aplicar (2.8) y (3.2):

$$1 = F(\infty) = \lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(X \leq x) = \lim_{x\to\infty} \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{\infty} f(t)\mathrm{d}t.$$

Este resultado se obtiene también al aplicar (3.5), con $B = \mathbf{R} = (-\infty, \infty)$.
($\Leftarrow$). Recíprocamente, si una función real $f(x)$ satisface (3.7)-(3.8), defina la función $F(x)$ como el lado derecho de (3.2). Esta es una función continua respecto de x. Además, $F(x)$ satisface las tres propiedades del Teorema 2.1.3. Por lo tanto, $F(x)$ corresponde a la función de distribución de cierta variable aleatoria X, la cual es continua. □

3.3. Distribuciones de probabilidad continuas

Esta sección describe los modelos paramétricos más usados para variables aleatorias continuas. Los temas de esta sección se desarrollan también en las siguientes dos secciones de este capítulo: 3.4 y 3.6. En ese sentido, se pretende conocer las propiedades básicas de las distribuciones de probabilidad continuas paramétricas: exponencial, uniforme continua, normal, gamma, beta, Weibull y Cauchy.

Distribución exponencial. En esta sección se define la distribución exponencial. Ejemplo que se aprovecha como introducción a las distribuciones de probabilidad continuas.

El tiempo de atención en un servidor, la vida útil de un foco, la vida útil de la estructura de un edificio, el tiempo de espera a la llegada a la tierra de un meteorito de al menos 500 kg de masa, son ejemplos de variable aleatorias de distribución exponencial. El conjunto de valores posibles de esta variable aleatoria es $X > 0$. Formalmente, una variable aleatoria X tiene *distribución exponencial* si su función de distribución es

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0, \end{cases} \tag{3.9}$$

donde $\lambda > 0$ es un parámetro *tasa*, que se considera fijo. La tasa indica la velocidad con que ocurren los eventos del fenómeno de estudio, por unidad de tiempo, longitud u otra unidad de medida. Por ejemplo, la tasa de llegada a la tierra de meteoritos de peso mínimo de 5 kg es mucho mayor que la tasa correspondiente de meteoritos con peso mínimo de 500 kg. Note que la función de distribución es continua, pues lo es para cada $x \in \mathbf{R}$. Además $F(x)$ es diferenciable para $x \neq 0$. La respectiva función de densidad es

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x) = F'(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x < 0. \end{cases} \tag{3.10}$$

Esta expresión se simplifica como

$$f(x) = \lambda e^{-\lambda x}; \quad x > 0, \quad \lambda > 0. \tag{3.11}$$

Con esta representación se destaca el conjunto de valores posibles de la variable aleatoria en estudio X:

$$\mathcal{X} = \{x \in \mathbf{R} : f(x) > 0\} = (0, \infty).$$

Así mismo, se destaca el *espacio paramétrico* $\lambda > 0$, que es la región factible de el *parámetro* o los parámetros del modelo, los cuales se consideran fijos o dados. La función de distribución (3.9) se simplifica como

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}; \quad x > 0, \quad \lambda > 0.$$

La función densidad $f(x)$ satisface las propiedades características (3.7)-(3.8), pues es no negativa e integrable, con

$$\begin{aligned} \int_{-\infty}^{\infty} f(t)\mathrm{d}t &= \int_{-\infty}^{0} f(t)\mathrm{d}t + \int_{0}^{\infty} f(t)\mathrm{d}t \\ &= \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{\infty} \lambda e^{-\lambda t}\mathrm{d}t \\ &= 0 + \int_{0}^{\infty} e^{-\lambda t}(\lambda\mathrm{d}t) = \int_{0}^{\infty} e^{-u}\mathrm{d}u \\ &= [-e^{-u}]_0^{\infty} = -[e^{-\infty} - e^{0}] \\ &= 1. \end{aligned} \tag{3.12}$$

Note que el cambio de variable $u = \lambda t$, simplificó el integrando a uno que corresponde a la función de densidad exponencial de tasa uno. De la misma manera, al integrar (3.11), se obtiene (3.9):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{x} 0\mathrm{d}t = 0, \quad \text{para} \quad x \leq 0$$

y

$$\begin{aligned}
F(x) = P(X \leq x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{0} f(t)\mathrm{d}t + \int_{0}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{0} 0\mathrm{d}t + \int_{0}^{x} \lambda e^{-\lambda t}\mathrm{d}t = \int_{0}^{x} e^{-\lambda t}(\lambda \mathrm{d}t) \\
&= \int_{0}^{\lambda x} e^{-u}\mathrm{d}u = [-e^{-u}]_{0}^{\lambda x} \\
&= -[e^{-\lambda x} - e^{0}] \\
&= 1 - e^{-\lambda x}; \quad \text{para} \quad x \geq 0.
\end{aligned}$$

De nuevo, con el cambio de variable $u = \lambda t$, el último integrando corresponde a la función de densidad exponencial de tasa uno.

3.4. Esperanza

Similar al caso discreto, las variables aleatorias continuas tienen conceptos asociados que caracterizan la forma de su distribución, como esperanza, varianza, momentos, función generadora de momentos, entre otros. En este caso, la sumatoria se sustituye por la integral.

Sea X una variable aleatoria continua con función de densidad $f(x) = F'(x)$; para $x \in \mathbf{R}$. Si

$$\int_{-\infty}^{\infty} |x| f(x)\mathrm{d}x < \infty,$$

su *esperanza* o *media poblacional* es

$$\mu = \mathrm{E}\, X = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x. \tag{3.13}$$

La esperanza es el centro de una población, de modo que sus individuos se encuentran al rededor de ella.

Ejemplo 3.4.1. Para una población de distribución exponencial de tasa $\lambda > 0$, su esperanza es

$$\mu = \mathrm{E}\, X = \frac{1}{\lambda}. \tag{3.14}$$

De hecho

$$\begin{aligned}
\mu = \mathrm{E}\, X &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} x0\mathrm{d}x + \int_{0}^{\infty} x\lambda e^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}\int_{0}^{\infty} (\lambda x) e^{-\lambda x}(\lambda \mathrm{d}x) \\
&= \frac{1}{\lambda}\int_{0}^{\infty} t e^{-t}\mathrm{d}t.
\end{aligned}$$

El cambio de variable $t = \lambda x$, reduce la integral a la esperanza de una variable aleatoria de distribución exponencial de tasa uno. Al considerar la técnica de integración por partes, con $u = t$ y $\mathrm{d}v = e^{-t}\mathrm{d}t$, se tiene

$$\mathrm{d}u = \mathrm{d}t, \quad v = \int e^{-t}\mathrm{d}t = -e^{-t}$$

e

$$\int_0^\infty te^{-t}\mathrm{d}t = [t(-e^{-t})]_0^\infty - \int_0^\infty [-e^{-t}]\mathrm{d}t = 0 + \int_0^\infty e^{-t}\mathrm{d}t = 1. \tag{3.15}$$

La última igualdad se debe a la propiedad característica de las densidades continuas (3.8), pues el último integrando corresponde a la función de densidad de distribución exponencial de tasa uno: $\lambda = 1$. Véase también expresión (3.12). Por otro lado, de la regla de L'Hôpital, se tiene

$$\lim_{t\to\infty} \frac{t}{e^t} = \frac{1}{\lim_{t\to\infty} e^t} = 0.$$

Por lo tanto, se deduce (3.14).

La media poblacional μ es también un parámetro de *escala*. En general, todo parámetro de escala es el recíproco de otro de tasa, y viceversa. En este sentido, la función de densidad (3.11) se escribe también como

$$f(x) = \frac{1}{\mu}e^{-x/\mu}; \quad x > 0, \quad \mu > 0. \tag{3.16}$$

Por ejemplo, considere dos servidores con tiempos de servicio de tasas $\lambda = 5.6$ y $\lambda = 8.2$ clientes por hora; respectivamente. El segundo servidor tiene una mayor tasa de servicio. Al atender más clientes por hora, el tiempo medio de servicio es menor. En la Figura ***[izquierda] se muestra la gráfica de la función de distribución exponencial, para ambos valores del parámetro tasa, junto con sus respectivas funciones de densidad [derecha]. Como

$$\lambda_1 = 5.6 \text{ clientes/h} < 8.2 \text{ clientes/h} = \lambda_2,$$

entonces

$$\mathrm{E}\,X_1 = \mu_1 = \frac{1}{\lambda_1} = \frac{1}{5.6} = 0.1786 \text{ h} > 0.1220 \text{ h} = \frac{1}{8.2} = \frac{1}{\lambda_2} = \mu_2 = \mathrm{E}\,X_2.$$

Asuma que la vida útil de una pala de acero es de distribución exponencial de tasa $\lambda = 2/\text{año}$, $X \sim \exp(\lambda = 2)$. Por desgaste o daño natural, se requieren dos palas al año; o bien, una pala cada medio año (seis meses). La esperanza de la variable aleatoria X es

$$\mu = \mathrm{E}\,X = \frac{1}{2} = 0.5 \text{ año}.$$

Por (3.9) y (3.11), las funciones de distribución y de densidad son respectivamente

$$F(x) = P(X \leq x) = 1 - e^{-2x}; \quad x > 0,$$

y

$$f(x) = F'(x) = 2e^{-2x}; \quad x > 0.$$

La probabilidad de que la pala dure un año o menos es

$$P(X \leq 1) = F(1) = 1 - e^{-2(1)} = 1 - e^{-2} = 1 - 0.1353 = 0.8647.$$

Su complemento, que la pala sobreviva un año, tiene probabilidad

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - (1 - 0.1353) = 0.1353.$$

Con la fórmula (3.4), para $0 \leq a \leq b \leq \infty$, se obtiene

$$\begin{aligned} P(a < X \leq b) = F(b) - F(a) &= \int_a^b f(t)\mathrm{d}t \\ &= \int_a^b \lambda e^{-\lambda t}\mathrm{d}t = \int_{\lambda a}^{\lambda b} e^{-u}\mathrm{d}u \\ &= [-e^{-u}]_{\lambda a}^{\lambda b} = -[e^{-\lambda b} - e^{-\lambda a}] \\ &= e^{-\lambda a} - e^{-\lambda b}. \end{aligned}$$

Este resultado se deduce también por (3.9):

$$P(a < X \leq b) = F(b) - F(a) = [1 - e^{-\lambda b}] - [1 - e^{-\lambda a}] = e^{-\lambda a} - e^{-\lambda b}.$$

En particular

$$\begin{aligned} P(0.5 < X \leq 1.5) &= F(1.5) - F(0.5) = [1 - e^{-2(1.5)}] - [1 - e^{-2(0,5)}] \\ &= e^{-1} - e^{-3} = 0.3679 - 0.0498 \\ &= 0.3181. \end{aligned}$$

Si la empresa ferretera ofrece una garantía de un mes por defectos de la pala, ¿qué porcentaje de clientes reclamará la garantía? Tanto para el cliente como el vendedor, el evento de interés es $\{0 < X \leq 1/12\}$. Así que

$$P\left(X \leq \frac{1}{12}\right) = F\left(\frac{1}{12}\right) = 1 - e^{-2(1/12)} = 1 - e^{-1/6} = 1 - 0.8464 = 0.1535.$$

El 15.35 % de los clientes recurrirán a la garantía. El resto de los clientes no podrán reclamar la garantía ante un daño posterior a un mes de uso. La calidad de la pala se resume con el valor del parámetro tasa: $\lambda = 2$. Con una menor tasa, la vida útil de la pala será mayor. Por ejemplo, considere que otro proveedor de palas ofrece una tasa λ, de modo que recibe un 10 % de reclamaciones durante el mismo período de garantía de un mes:

$$\begin{aligned} 0.1 = P\left(X \leq \frac{1}{12}\right) &= F\left(\frac{1}{12}\right) \\ &= 1 - e^{-\lambda(1/12)} = 1 - e^{-\lambda/12}. \end{aligned}$$

Como la pala es de mejor calidad, la tasa será menor a 2:

$$
\begin{aligned}
e^{-\lambda/12} &= 1 - 0.1 = 0.9, \\
-\frac{\lambda}{12} &= \log 0.9 = -0.1053, \\
\frac{\lambda}{12} &= 0.1053, \\
\lambda &= 12(0.1053) = 1.2643 < 2.
\end{aligned}
$$

Al mejorar la tasa, la pala del nuevo proveedor tiene una vida media mayor:

$$\mu = \frac{1}{\lambda} = \frac{1}{1.2643} = 0.7909\,\text{año} > 0.5\,\text{año}.$$

Este ejemplo desarrollado de la distribución exponencial muestra la utilidad de caracterizar a una población en términos de un modelo de distribución de probabilidad. En consecuencia, los modelos de distribuciones de probabilidad resuelven situaciones prácticas como:

1. Planificación de presupuestos de una compañía constructora
2. Control de inventarios
3. Política de mantenimiento y renovación de maquinaria y equipo
4. Determinación del precio de la pala
5. Determinación del período de garantía
6. Especificación de la calidad de la pala por el parámetro tasa λ, o por su equivalente media $\mu = 1/\lambda$; basado en el modelo de probabilidad de distribución exponencial
7. Regulación en temas de seguridad e higiene.

Ahora se describirá la fórmula de momentos para la transformación de una variable aleatoria continua. Sea $X : \Omega \to \mathcal{X}$, una variable aleatoria continua, con conjunto de valores posibles $\mathcal{X} \subset \mathbf{R}$ y función de densidad $f(x) = F'(x)$; para $x \in \mathbf{R}$. Considere la transformación $Y = g(X)$, donde $g : \mathcal{X} \to \mathbf{R}$, es una función real continua definida en $\mathcal{X} \subset \mathbf{R}$. Entonces, Y también es una variable aleatoria. Si además se satisface

$$\int_{-\infty}^{\infty} |g(x)| f(x) \mathrm{d}x < \infty,$$

entonces

$$\mathrm{E}\, Y = \mathrm{E}\, g(X) = \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x. \tag{3.17}$$

Por ejemplo

$$\mathrm{E}\, X^2 = \int_{-\infty}^{\infty} x^2 f(x) \mathrm{d}x,$$

siempre que dicha integral sea finita. En tal caso, la *varianza* de la variable aleatoria X es

$$\sigma^2 = \operatorname{V} X = \operatorname{E}(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x,$$

donde $\mu = \operatorname{E} X$. Así mismo, la medida de dispersión *desviación estándar* se define como en (2.30):

$$\sigma = \mathrm{d.\,e.}(X) = \sqrt{\operatorname{V} X} = \sqrt{\sigma^2}.$$

Las propiedades de la esperanza y varianza, descritas en los teoremas 2.4.4 y 2.5.1, aplican también para variables aleatorias continuas. En particular

$$\sigma^2 = \operatorname{V} X = \operatorname{E}(X-\mu)^2 = \operatorname{E} X^2 - \mu^2.$$

De hecho

$$\begin{aligned}
\operatorname{V} X = \operatorname{E}(X-\mu)^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x + \int_{-\infty}^{\infty} (-2\mu x) f(x)\mathrm{d}x + \int_{-\infty}^{\infty} \mu^2 f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x - 2\mu \int_{-\infty}^{\infty} x f(x)\mathrm{d}x + \mu^2 \int_{-\infty}^{\infty} f(x)\mathrm{d}x \\
&= \operatorname{E} X^2 - 2\mu \operatorname{E} X + \mu^2 \\
&= \operatorname{E} X^2 - 2\mu(\operatorname{E} X) + (\operatorname{E} X)^2 \\
&= \operatorname{E} X^2 - (\operatorname{E} X)^2.
\end{aligned}$$

La *función generadora de momentos* es

$$M(t) = \operatorname{E} e^{tX} = \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x; \quad \text{para} \quad t \in \mathbf{R}. \tag{3.18}$$

Como en el caso discreto, esta función satisface (2.67) y (2.68):

$$M^{(k)}(t) = \operatorname{E}[X^k e^{tX}]; \quad t \in \mathbf{R}, \quad k = 1, 2, \ldots$$

y

$$\operatorname{E} X^k = M^{(k)}(0); \quad k = 1, 2, \ldots$$

En particular, se verifica

$$M'(t) = \frac{\mathrm{d}}{\mathrm{d}t} \operatorname{E} e^{tX} = \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{\partial}{\partial t}\{e^{tx} f(x)\}\mathrm{d}x = \int_{-\infty}^{\infty} x e^{tx} f(x)\mathrm{d}x = \operatorname{E}[X e^{tX}]$$

y

$$\begin{aligned}
M''(t) = \frac{\mathrm{d}}{\mathrm{d}t} \operatorname{E}[X e^{tX}] &= \frac{\mathrm{d}}{\mathrm{d}t} \int_{-\infty}^{\infty} x e^{tx} f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial t}\{e^{tx} x f(x)\}\mathrm{d}x = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\mathrm{d}x \\
&= \operatorname{E}[X^2 e^{tX}].
\end{aligned}$$

Así que

$$M(0) = \mathrm{E}\, e^{0X} = 1,$$
$$M'(0) = \mathrm{E}[Xe^{0X}] = \mathrm{E}\, X,$$
$$M''(0) = \mathrm{E}[X^2 e^{0X}] = \mathrm{E}\, X^2.$$

Para una variable aleatoria X de distribución exponencial de tasa $\lambda > 0$ (media $\mu = 1/\lambda > 0$), se tiene

$$\sigma^2 = \mathrm{V}\, X = \mu^2 = \frac{1}{\lambda^2}, \tag{3.19}$$

$$\sigma = \sqrt{\mathrm{V}\, X} = \mu = \frac{1}{\lambda} \tag{3.20}$$

y

$$M(t) = \mathrm{E}\, e^{tX} = \frac{1}{1-\mu t} = \frac{\lambda}{\lambda - t}; \quad \text{para} \quad t < \lambda = \frac{1}{\mu}. \tag{3.21}$$

Por lo cual, la varianza de una población de distribución exponencial es grande si su media poblacional es también lo es. Note además que su función generadora de momentos no está definida para $t \geq \lambda$, pues la integral correspondiente es ∞. Se verificará (3.19)-(3.21). Considere

$$\begin{aligned}\mathrm{E}\, X^2 &= \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = \int_{-\infty}^{0} x^2 0\mathrm{d}x + \int_{0}^{\infty} x^2 \lambda e^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\lambda^2}\int_0^{\infty} (\lambda x)^2 e^{-\lambda x}(\lambda \mathrm{d}x) = \mu^2 \int_0^{\infty} t^2 e^{-t}\mathrm{d}t.\end{aligned}$$

Por la técnica de integración por partes, con $u = t^2$ y $\mathrm{d}v = e^{-t}\mathrm{d}t$, se obtiene

$$\mathrm{d}u = 2t\mathrm{d}t, \quad v = \int e^{-t}\mathrm{d}t = -e^{-t}$$

e

$$\int_0^{\infty} t^2 e^{-t}\mathrm{d}t = [t^2(-e^{-t})]_0^{\infty} - \int_0^{\infty} [-e^{-t}](2t\mathrm{d}t) = 2\int_0^{\infty} te^{-t}\mathrm{d}t = 2 \times 1 = 2.$$

La última integral es la esperanza de una variable aleatoria de distribución exponencial de tasa uno; véase (3.14) o su deducción en (3.15). Por otro lado, de la regla de L'Hôpital, se tiene

$$\lim_{t\to\infty} \frac{t^2}{e^t} = \lim_{t\to\infty} \frac{2t}{e^t} = \frac{2}{\lim_{t\to\infty} e^t} = 0.$$

Por lo tanto

$$\mathrm{E}\, X^2 = 2\mu^2 = \frac{2}{\lambda^2} \tag{3.22}$$

y

$$\mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = 2\mu^2 - \mu^2 = \mu^2 = \frac{1}{\lambda^2}.$$

Por otro lado, la función generadora de momentos se obtiene por lo siguiente

$$\begin{aligned}
M(t) = \mathrm{E}\, e^{tX} &= \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x \\
&= \int_{-\infty}^{0} e^{tx} 0\mathrm{d}x + \int_{0}^{\infty} e^{tx}\lambda e^{-\lambda x}\mathrm{d}x = \lambda \int_{0}^{\infty} e^{tx-\lambda x}\mathrm{d}x \\
&= \lambda \int_{0}^{\infty} e^{-(\lambda-t)x}\mathrm{d}x = \frac{\lambda}{\lambda - t}\int_{0}^{\infty} (\lambda - t)e^{-(\lambda-t)x}\mathrm{d}x \\
&= \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda} \\
&= \frac{1}{1-\mu t}.
\end{aligned}$$

El último integrando corresponde a la función de densidad exponencial de tasa $\lambda - t > 0$. Por lo cual integra uno. En cambio, si $\lambda - t \leq 0$, dicha integral es $-\infty$. Se concluye (3.21). Ahora se deducirán los primeros dos momentos de la distribución exponencial, vía su función generadora. Por (3.21), se tiene

$$\begin{aligned}
M'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{1-\mu t}\right) = \frac{\mathrm{d}}{\mathrm{d}t}(1-\mu t)^{-1} = -(-\mu)(1-\mu t)^{-2} = \frac{\mu}{(1-\mu t)^2}, \\
M''(t) &= \mu \frac{\mathrm{d}}{\mathrm{d}t}(1-\mu t)^{-2} = -2\mu(-\mu)(1-\mu t)^{-3} = \frac{2\mu^2}{(1-\mu t)^3}, \\
M^{(3)}(t) &= 2\mu^2 \frac{\mathrm{d}}{\mathrm{d}t}(1-\mu t)^{-3} = -2(3)\mu^2(-\mu)(1-\mu t)^{-4} = \frac{6\mu^3}{(1-\mu t)^4}.
\end{aligned}$$

Por inducción matemática, se obtiene

$$M^{(k)}(t) = \frac{k!\mu^k}{(1-\mu t)^{k+1}}; \quad \text{para} \quad t < 1/\mu, \quad k = 0, 1, 2, \ldots.$$

Al evaluar estas derivadas en $t = 0$, se tiene

$$EX^k = M^{(k)}(0) = \frac{k!\mu^k}{(1-\mu \times 0)^{k+1}} = k!\mu^k; \quad \text{para} \quad k = 0, 1, 2, \ldots.$$

En particular

$$\begin{aligned}
EX &= M'(0) = \mu, \\
EX^2 &= M''(0) = 2\mu^2, \\
EX^3 &= M^{(3)}(0) = 6\mu^3, \\
EX^4 &= M^{(4)}(0) = 24\mu^4.
\end{aligned}$$

Las primeras dos igualdades de la expresión anterior confirman las fórmulas de momentos (3.14) y (3.22). Las otras dos igualdades se son útiles para el cálculo de los coeficientes de sesgo y curtosis. Por (2.77), resulta

$$\begin{aligned}
\mathrm{E}(X-\mu)^3 &= \mathrm{E}\, X^3 - 3\mu\, \mathrm{E}\, X^2 + 2\mu^3 \\
&= 6\mu^3 - 3\mu(2\mu^2) + 2\mu^3 \\
&= 2\mu^3
\end{aligned}$$

y

$$\text{sesgo}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\mathrm{E}(X-\mu)^3}{\mu^3} = \frac{2\mu^3}{\mu^3} = 2.$$

Por (3.20), recuerde que la desviación estándar de la distribución exponencial es la media poblacional μ. De manera similar, se obtiene el coeficiente de curtosis. Para ello, considere (2.78):

$$\begin{aligned}\mathrm{E}(X-\mu)^4 &= \mathrm{E}\,X^4 - 4\mu\,\mathrm{E}\,X^3 + 6\mu^2\,\mathrm{E}\,X^2 - 3\mu^4\\ &= 24\mu^4 - 4\mu(6\mu^3) + 6\mu^2(2\mu^2) - 3\mu^4\\ &= 9\mu^4 > 0\end{aligned}$$

y

$$\text{curtosis}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^4 = \frac{\mathrm{E}(X-\mu)^4}{\mu^4} = \frac{9\mu^4}{\mu^4} = 9.$$

Como el sesgo es positivo y la curtosis es relativamente grande, entonces la población es asimétrica con cola derecha pesada. Note además que, tanto el sesgo como la curtosis de la distribución exponencial no dependen del parámetro tasa. De hecho, la forma de la función de densidad es la misma, para cualquier tasa $\lambda > 0$. Véanse las gráficas de las funciones de densidad de la Figura 3.1[arriba].

Por ejemplo, considere que los tiempos entre llegada de clientes a una gasolinera tiene distribución exponencial de media poblacional

$$\mu = 3\text{min}.$$

Entonces, la tasa es

$$\lambda = \frac{1}{\mu} = \frac{1}{3} = 0.3333\,/\text{min}.$$

En la Figura 3.1 se muestra la gráfica de las funciones de densidad exponencial, con parámetro media $\mu = 0.5, 1, 3$. Al incrementar el parámetro μ, la forma de la función de densidad es cada vez más plana o llana, causando una mayor dispersión de la población. El cero es la moda, más sin embargo este no es un valor posible de X. Note que, en el primer ejemplo, la función de densidad $f(x)$ es mayor que uno, pues $f(x)$ no es una probabilidad. Incluso, hay algunos otros ejemplos donde $f(x)$ no está acotada superiormente.

Por el Ejercicio 3.3, la distribución exponencial no tiene memoria, en el sentido de que "nuevo es igual que usado":

$$P(X > x + y \mid X > x) = P(X > y); \quad x, y \geq 0.$$

Esta expresión equivale a

$$P(X \leq x + y \mid X > x) = P(X \leq y); \quad x, y \geq 0.$$

Por ejemplo, asuma que la vida útil de un foco tiene distribución exponencial de media $\mu = 2.5$ años. Entonces, la probabilidad de que un foco seleccionado al azar dure más de 2

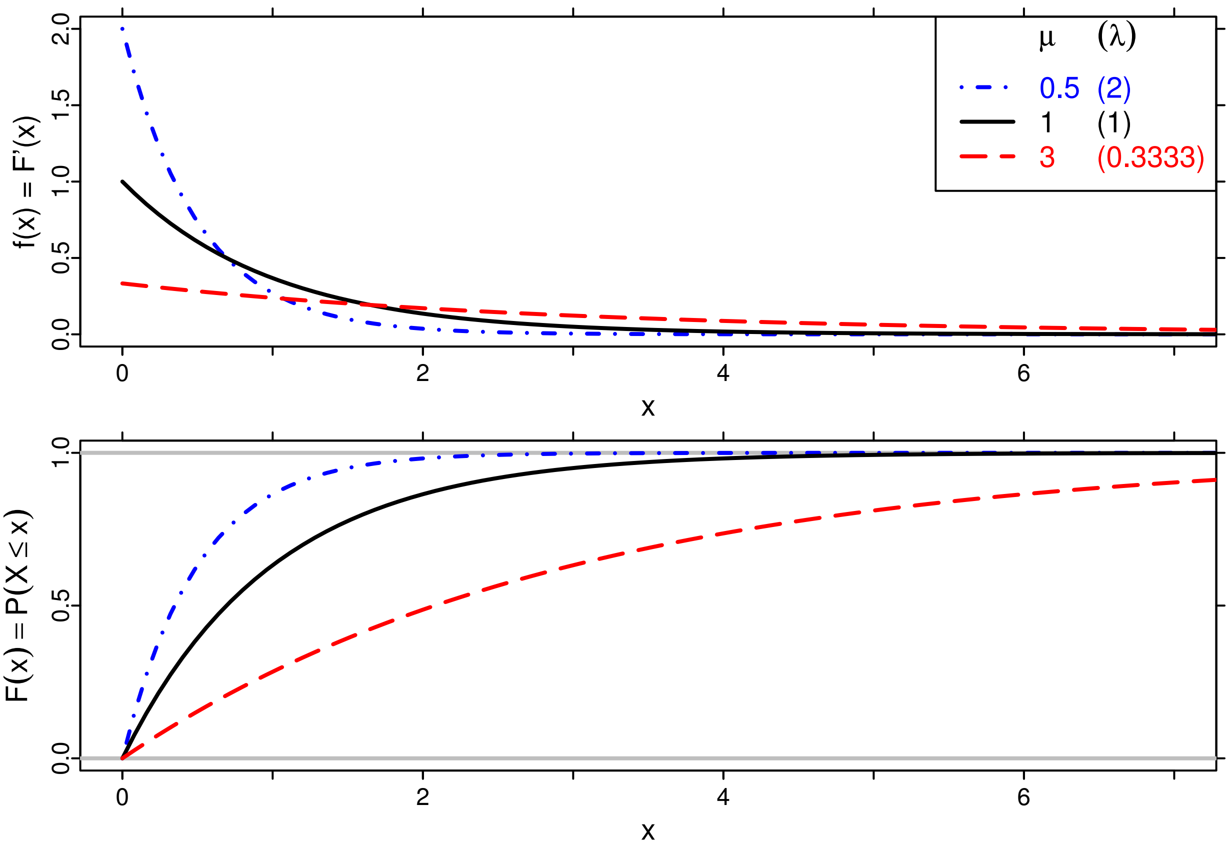


Figura 3.1. [arriba] Gráficas de la función de densidad exponencial de medias $\mu = 0.5, 1, 3$. [abajo] Correspondientes gráficas de la función de distribución.

años es

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - F(2) \\ &= 1 - (1 - e^{-2/\mu}) = 1 - (1 - e^{-2/2.5}) \\ &= e^{-0.8} = 0.4493. \end{aligned}$$

Ahora bien, si el foco ya tiene una edad de 3.5 años, entonces este está como nuevo, y

$$P(X > 5.5 \mid X > 3.5) = P(X > 3.5 + 2 \mid X > 3.5) = P(X > 2) = 0.4493.$$

La familia de distribuciones exponencial es cerrada bajo transformaciones de escala:

$$X \sim \exp(\mu) \quad \Leftrightarrow \quad Y = bX \sim \exp(b\mu); \quad \text{con} \quad b > 0.$$

Lo que confirma que μ es un parámetro de escala. Para verificar lo anterior se puede usar la función generadora de momentos y sus propiedades. Recuerde el Teorema 2.6.8.

$$M_Y(t) = M_{bX}(t) = M_X(bt) = \frac{1}{1 - \mu(bt)} = \frac{1}{1 - (b\mu)t}.$$

Esta expresión corresponde a la función generadora de momentos de la distribución exponencial de media $b\mu$, la cual es finita para $t < 1/(b\mu)$. Otra vía de verificación es con el cálculo de la función de distribución de Y:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(bX \leq y) \\ &= P\left(X \leq \frac{y}{b}\right) = F_X\left(\frac{y}{b}\right) \\ &= 1 - e^{-(y/b)/\mu} = 1 - e^{-y/(b\mu)}; \quad y \geq 0. \end{aligned}$$

Esta expresión corresponde a la función de distribución exponencial de media $b\mu$ (3.9). Su función de densidad es

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}[1 - e^{-y/(b\mu)}] = \frac{1}{b\mu} e^{-y/(b\mu)}; \quad y > 0.$$

Compare esta expresión con (3.11).

Ejemplo 3.4.2. Un director de Universidad elegido al azar tiene un ingreso mensual modelado bajo una distribución exponencial *trasladada*, de parámetros localización $a = \$118,500.00$ pesos y escala $\mu = \$5,000.00$ pesos. En este caso, la función de densidad es

$$f(x) = \frac{1}{\mu} e^{-(x-a)/\mu} = \frac{1}{5000} e^{-(x-118500)/5000}; \quad x > 118500.$$

Esta función es positiva a partir de $x = a = 118500$. Su forma es idéntica a la función de densidad exponencial, sólo que trasladada a unidades a la derecha. ¿Cuál es la probabilidad de que el ingreso de dicho funcionario sea mayor que $\$120,000.00$ pesos? Solución. Sea X

la variable aleatoria que representa el ingreso del funcionario. El evento de interés es $\{X > 120000\}$. Entonces

$$\begin{aligned} P(X > 120000) &= \int_{120000}^{\infty} \frac{1}{5000} e^{-(x-118500)/5000} \mathrm{d}x = \int_{(120000-118500)/5000}^{\infty} e^{-u} \mathrm{d}u \\ &= \int_{0.3}^{\infty} e^{-u} \mathrm{d}u = [-e^{-u}]_{0.3}^{\infty} \\ &= -[e^{-\infty} - e^{-0.3}] = e^{-0.3} \\ &= 0.7408. \end{aligned}$$

Aquí se usó el cambio de variable $u = (x-118500)/5000$. Otra solución se obtiene al considerar la transformación

$$Y = X - a = X - 118500,$$

de modo que Y tiene distribución exponencial de media $\mu = 5000$. Así que

$$\begin{aligned} P(X > 120000) &= P(X - 118500 > 120000 - 118500) \\ &= P(Y > 1500) \\ &= 1 - [1 - e^{-1500/5000}] \\ &= e^{-0.3} \\ &= 0.7408. \end{aligned}$$

Una transformación similar es

$$Y = \frac{X-a}{\mu} = \frac{X - 118500}{5000},$$

de modo que Y tiene distribución exponencial de media uno. Por lo que

$$\begin{aligned} P(X > 120000) &= P(X - 118500 > 120000 - 118500) \\ &= P\left(\frac{X - 118500}{5000} > \frac{120000 - 118500}{5000}\right) \\ &= P(Y > 0.3) = 1 - [1 - e^{-0.3}] \\ &= e^{-0.3} = 0.7408. \end{aligned}$$

3.5. Distribuciones normal y log normal

En esta sección se definirá la distribución normal. Se abordan sus principales propiedades. Se describe también la distribución log normal, así como la relación con su contraparte normal.

La distribución normal es el modelo de probabilidad de mayor aplicación en todas las áreas del conocimiento, ya sea directamente o a través de una transformación. Esta distribución de probabilidad es un referente de modelación de primera instancia para poblaciones cuyos individuos se dispersan de manera simétrica al rededor del centro. El origen de la distribución normal fue la modelación de los errores de los instrumentos de medición, así como de observaciones astronómicas.

Una variable aleatoria X tiene una *distribución normal*, con media $\mu \in \mathbf{R}$ y varianza $\sigma^2 > 0$, si su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2}; \quad \text{para} \quad -\infty < x < \infty. \tag{3.23}$$

Se denota como $X \sim \mathrm{N}(\mu, \sigma^2)$. La media poblacional μ, es un parámetro de localización mientras que de escala es la desviación estándar σ. Recuerde que tanto la desviación estándar como la varianza son indicadores de dispersión de una población. El caso particular de la distribución *normal estándar* es cuando $\mu = 0$ y $\sigma^2 = 1$:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; \quad z \in \mathbf{R}. \tag{3.24}$$

La correspondiente función de distribución normal estándar es

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)\mathrm{d}t = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\mathrm{d}t; \quad z \in \mathbf{R}. \tag{3.25}$$

Esta función se obtiene sólo por métodos numéricos o tablas de probabilidad, pues la integral involucrada no tiene solución cerrada. La gráfica de la función de densidad normal estándar, así como de la correspondiente función de distribución, se muestra en la Figura 3.2. Note que la población normal estándar es simétrica respecto de la media cero; véase Ejercicio 3.24. De hecho, toda distribución normal es simétrica respecto de su media μ. Por lo cual, este parámetro de localización es también la moda y mediana poblacional. Para el caso general normal, la función de distribución normal a (3.23) satisface

$$F(x) = P(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right); \quad x \in \mathbf{R}. \tag{3.26}$$

Esta fórmula se deduce como sigue

$$\begin{aligned} F(x) = P(X \leq x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\mathrm{d}t = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z \\ &= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z)\mathrm{d}z = \Phi\left(\frac{x-\mu}{\sigma}\right); \quad x \in \mathbf{R}. \end{aligned}$$

Aquí se usó el cambio de variable $z = (x-\mu)/\sigma$. La función $f(x)$ satisface las dos propiedades características de una función de densidad continua, pues es positiva e

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z = 1. \tag{3.27}$$

La última igualdad se verifica con el Ejercicio 3.25. De nuevo, se usó el cambio de variable $z = (x-\mu)/\sigma$.

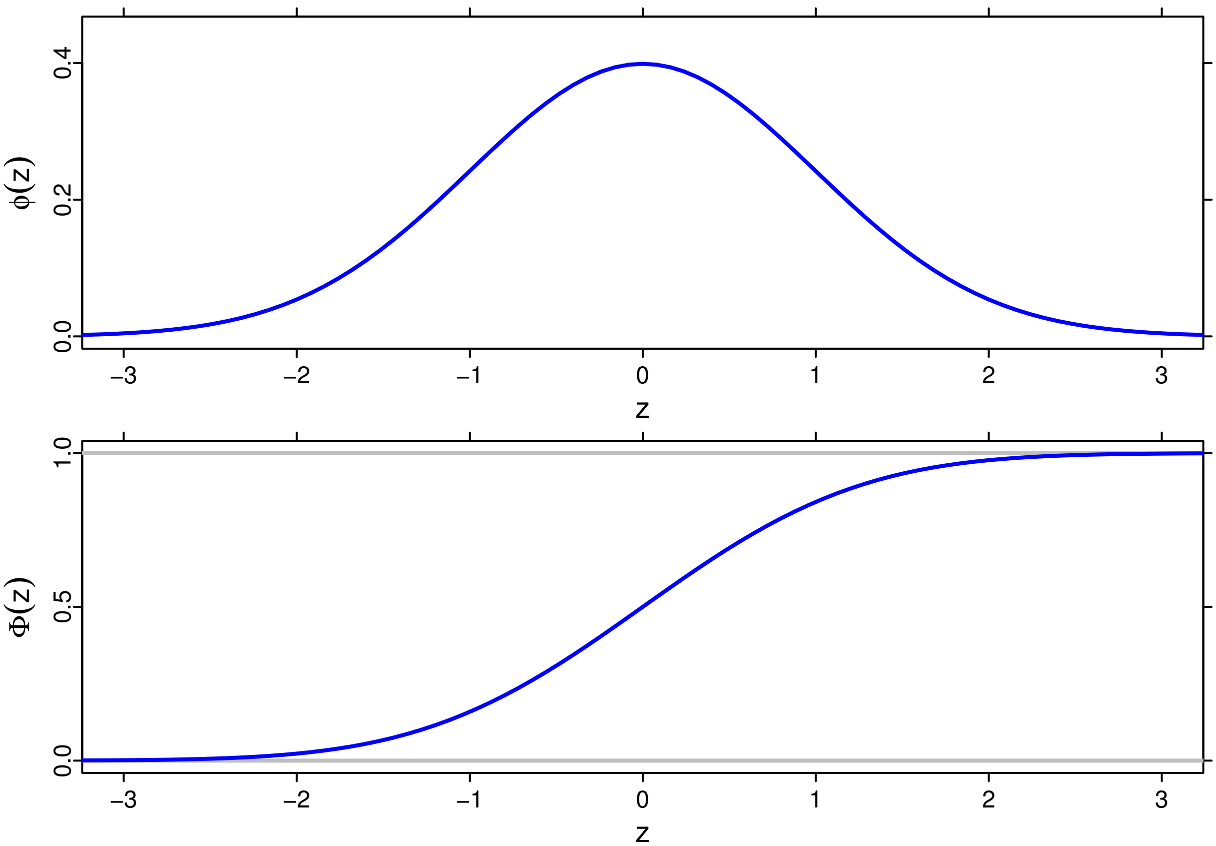


Figura 3.2. [arriba] Gráfica de la función de densidad normal estándar $\phi(z)$. [abajo] Correspondiente gráfica de la función de distribución $\Phi(z)$.

Ahora se deducirán la media y varianza de una población de distribución normal. Primero se desarrollará el caso normal estándar. Por (3.13), la esperanza de una variable aleatoria de distribución normal estándar Z es

$$\begin{aligned}
\mathrm{E}\, Z &= \int_{-\infty}^{\infty} z\phi(z)\mathrm{d}z = \int_{-\infty}^{\infty} z\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{0} e^{-z^2/2}(z\mathrm{d}z) + \int_{0}^{\infty} e^{-z^2/2}(z\mathrm{d}z)\right] \\
&= \frac{1}{\sqrt{2\pi}}\left[\int_{\infty}^{0} e^{-u}\mathrm{d}u + \int_{0}^{\infty} e^{-u}\mathrm{d}u\right] \\
&= \frac{1}{\sqrt{2\pi}}\left[-\int_{0}^{\infty} e^{-u}\mathrm{d}u + \int_{0}^{\infty} e^{-u}\mathrm{d}u\right] \\
&= \frac{1}{\sqrt{2\pi}}[-1+1] = 0.
\end{aligned}$$

Aquí se usó el cambio de variable $u = z^2/2$ para ambas integrales. Este resultado se verifica también por el hecho de que el integrando $z\phi(z)$ es una función *impar* o *anti simétrica* respecto de cero. Por lo que se anula su integral. De manera similar se deduce el segundo momento no central de Z:

$$\mathrm{E}\, Z^2 = \int_{-\infty}^{\infty} z^2\phi(z)\mathrm{d}z = \int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z = \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} ze^{-z^2/2}(z\mathrm{d}z).$$

La última igualdad se debe a que el integrando $z^2\phi(z)$ es una función *par* o *simétrica* respecto de cero. Considere la técnica de integración por partes, con $u = z$ y $\mathrm{d}v = e^{-z^2/2}(z\mathrm{d}z)$:

$$\mathrm{d}u = \mathrm{d}z, \quad v = \int e^{-z^2/2}(z\mathrm{d}z) = -e^{-z^2/2}$$

e

$$\int_{0}^{\infty} ze^{-z^2/2}(z\mathrm{d}z) = \left[-ze^{-z^2/2}\right]_{0}^{\infty} - \int_{0}^{\infty}[-e^{-z^2/2}]\mathrm{d}z = \int_{0}^{\infty} e^{-z^2/2}\mathrm{d}z.$$

Por lo que

$$\mathrm{E}\, Z^2 = \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-z^2/2}\mathrm{d}z = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z = \int_{-\infty}^{\infty} \phi(z)\mathrm{d}z = 1.$$

Se concluye

$$\mathrm{V}\, Z = \mathrm{E}\, Z^2 - (\mathrm{E}\, Z)^2 = 1 - 0 = 1.$$

Para el caso general de la distribución normal, debe comprobarse

$$\mathrm{E}\, X = \mu \quad \text{y} \quad \mathrm{V}\, X = \sigma^2.$$

Aquí se desarrollarán dos alternativas de solución. La primera implica el uso de un cambio de variable, como en (3.27). Para luego reducir la esperanza y varianza de X a la del caso normal estándar. Considere el cambio de variable

$$z = \frac{x-\mu}{\sigma}, \tag{3.28}$$

Entonces $x = \mu + \sigma z$, $\mathrm{d}z = \mathrm{d}x/\sigma$ y

$$\begin{aligned}
\mathrm{E}\,X &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{\mu+\sigma z}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z + \sigma \int_{-\infty}^{\infty} \frac{z}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z \\
&= \mu(1) + \sigma(0) = \mu.
\end{aligned}$$

Así mismo

$$\begin{aligned}
\mathrm{V}\,X = \mathrm{E}(X-\mu)^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x = \sigma^2 \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left(\frac{\mathrm{d}x}{\sigma}\right) \\
&= \sigma^2 \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\mathrm{d}z = \sigma^2(1) \\
&= \sigma^2.
\end{aligned}$$

El cambio de variable (3.28), en el contexto de integral, tiene una interpretación de probabilidad. La siguiente transformación de localización y escala es una *estandarización* de la variable aleatoria X:

$$Z = \frac{X-\mu}{\sigma}, \tag{3.29}$$

La nueva variable aleatoria Z tiene también distribución normal, pues un cambio de unidad de medida no perturba la esencia de la distribución. De hecho, su distribución es normal estándar. En consecuencia, la variable aleatoria Z tiene media cero y varianza uno:

$$\mathrm{E}\,Z = \mathrm{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathrm{E}[X-\mu]}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

y

$$\mathrm{V}\,Z = \mathrm{V}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathrm{V}[X-\mu]}{\sigma^2} = \frac{\mathrm{V}\,X}{\sigma^2} = \frac{\sigma^2}{\sigma^2}. = 1$$

La función de densidad normal estándar de la variable Z se deduce por lo siguiente.

$$\begin{aligned}
\Phi(z) = P(Z \le z) &= P\left(\frac{X-\mu}{\sigma} \le z\right) \\
&= P(X \le \mu + \sigma z) = F_X(\mu + \sigma z) \\
&= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x; \quad z \in \mathbf{R}.
\end{aligned}$$

Al considerar el cambio de variable $t = (x-\mu)/\sigma$, se tiene

$$t(-\infty) = -\infty, \quad t(\mu+\sigma z) = z, \quad \mathrm{d}t = \frac{\mathrm{d}x}{\sigma}$$

y

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t; \quad z \in \mathbf{R}.$$

Por lo tanto

$$Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1). \tag{3.30}$$

Recíprocamente, si Z es una variable aleatoria de distribución normal estándar $Z \sim \mathrm{N}(0, 1)$, entonces la variable aleatoria $X = \mu + \sigma Z$, tiene distribución normal de media μ y varianza σ^2. Se concluye

$$X = \mu + \sigma Z \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1).$$

En general, la familia de distribuciones normal es cerrada bajo transformaciones de localización y escala:

$$X \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Y = a + bX \sim \mathrm{N}(a + b\mu, b^2\sigma^2); \quad \text{con} \quad a \in \mathbf{R} \quad \text{y} \quad b > 0. \tag{3.31}$$

Este resultado confirma el rol de parámetro de localización de μ, así como el de escala de σ. Por otro lado, la equivalencia (3.31) aplica también para $b < 0$. En tal caso, la transformación implica un cambio del sentido de la unidad de medida. En particular

$$Z \sim \mathrm{N}(0, 1) \quad \Leftrightarrow \quad -Z \sim \mathrm{N}(0, 1). \tag{3.32}$$

Considere el siguiente ejemplo. Durante el año 2023, el salario registrado ante el Instituto Mexicano del Seguro social (IMSS) tuvo un incremento nominal de 10.5 %, mientras que la inflación en tal período de tiempo fue de 4.66 %; véase [4] y [7]. Sea la variable aleatoria X el incremento porcentual del ingreso de un trabajador que cotiza al IMSS, la cual se asume que tiene una distribución normal de media $\mu = 10.5\,\%$ y varianza hipotética $\sigma^2 = 12.8164$. La desviación estándar es $\sigma = \sqrt{\sigma} = \sqrt{12.8164} = 3.58\,\%$. Entonces

$$X \sim \mathrm{N}(\mu = 10.5, \sigma^2 = 12.8164).$$

EL modelo de distribución normal establece incluso decrementos nominales del ingreso ante el IMSS. Un intervalo de 95 % de predicción del incremento porcentual del ingreso es

$$\begin{aligned} X &\in \mu \pm 1.96\sigma \\ &= 10.5 \pm 1.96 \times 3.58 \\ &= 10.5 \pm 7.015 \\ &= (3.4833\,\%, 17.5167\,\%). \end{aligned}$$

La función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}3.58} e^{-\frac{1}{2}\left(\frac{x-10.5}{3.58}\right)^2}; \quad x \in \mathbf{R}.$$

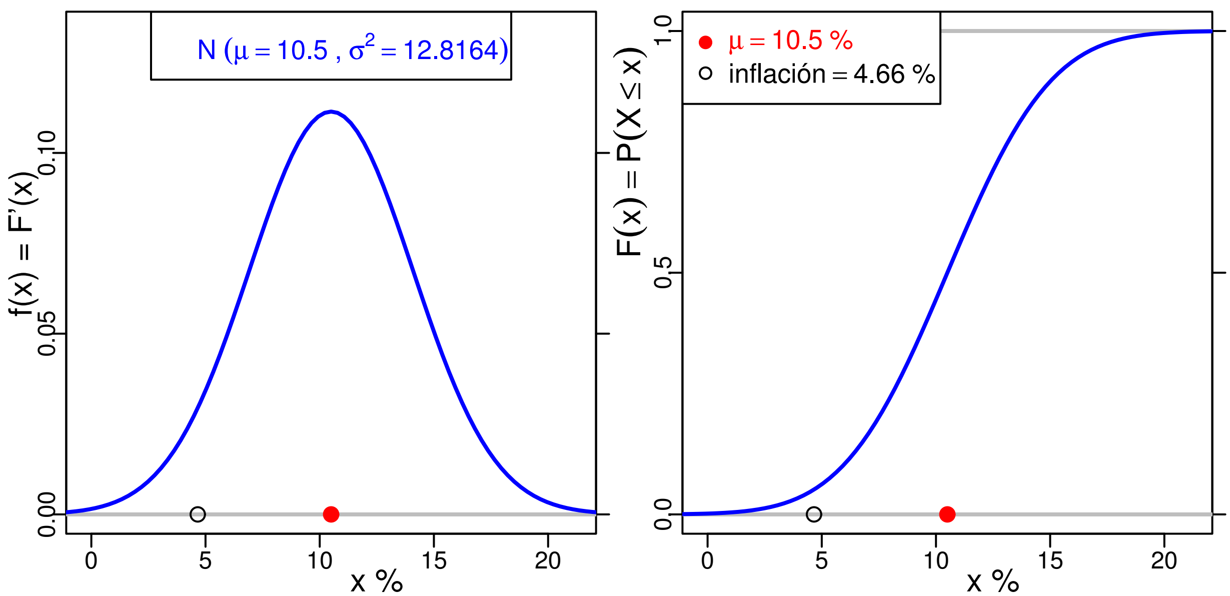


Figura 3.3. [izquierda] Gráfica de la función de densidad normal; modelo del incremento porcentual del ingreso del trabajador ante el IMSS durante 2023. [derecha] Respectiva gráfica de función de distribución.

La correspondiente función de distribución es

$$\begin{aligned} F(x) = P(X \leq x) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-10.5}{3.58}\right) = \Phi\left(\frac{x-10.5}{3.58}\right); \quad x \in \mathbf{R}. \end{aligned}$$

La Figura 3.3 muestra la gráfica de la función de densidad y la función de distribución de la variable aleatoria X. Por otro lado, al considerar que durante 2023 el índice de precios tuvo un incremento de 4.66 %, se aprecia que la mayoría de los trabajadores registrados ante el IMSS mejoraron sus ingresos en términos reales:

$$P(X > 4.66) = 1 - P(X \leq 4.66),$$

con

$$\{X \leq 4.66\} = \left\{\frac{X-\mu}{\sigma} \leq \frac{4.66-\mu}{\sigma}\right\} = \left\{Z \leq \frac{4.66-10.5}{3.58}\right\} = \{Z \leq -1.6313\}$$

y

$$P(X \leq 4.66) = P(Z \leq -1.6313) = \Phi(-1.6313) = 0.0514 = 5.14\,\%.$$

Así

$$P(X > 4.66) = 1 - 0.0514 = 0.9486 = 94.86\,\%.$$

levantamiento	mujeres	hombres	brecha salarial %
2016	16,097	27,875	42.25
2022	19,081	29,285	34.84

Tabla 3.1. Ingreso promedio trimestral en México, por sexo y año del levantamiento muestral; a pesos de 2022.

La Tabla 3.1 muestra el ingreso promedio trimestral de los mexicanos, por sexo y año de levantamiento, a pesos de 2022, según Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH); véase [6, Cuadro 10, p.p. 18]. Asuma que las cuatro poblaciones aludidas tienen una distribución normal de media la especificada en dicha tabla y desviación estándar común: $\sigma = 1,250$ pesos. Por ejemplo, si la variable aleatoria X es el ingreso trimestral de una trabajadora de 2022 seleccionada al azar, entonces

$$\mu = 19.081 \text{ mil pesos}, \quad \sigma = 1.25 \text{ mil pesos}, \quad \sigma^2 = 1.5625 \text{ (mil pesos)}^2$$

y

$$X \sim \mathrm{N}(\mu = 19.081, \sigma^2 = 1.5625).$$

La función de densidad de esta población es

$$f(x) = \frac{1}{\sqrt{2\pi}1.25} e^{-\frac{1}{2}\left(\frac{x-19.081}{1.25}\right)^2}; \quad x \in \mathbf{R}.$$

Su gráfica, junto con las de las otras tres poblaciones, se muestra en la Figura 3.4. Cada función de densidad refleja la dispersión de su población. Note que en términos reales, el ingreso a aumentado tanto para hombres como para mujeres, especialmente en este último grupo. Si bien en un sexenio se ha reducido la brecha salarial, aún es significativamente amplia.

Asuma que en un procedimiento industrial se desea mejorar la media poblacional. Para ello se establecen dos tratamientos: control e innovación. Considere que ambas poblaciones tienen distribución normal, de diferente media e igual varianza:

$$\mu_1 = 10, \quad \mu_2 = 16.5, \quad \sigma_1^2 = \sigma_2^2 = 2.$$

La Figura 3.5 muestra las gráficas de ambas funciones de densidad. Aquí se describe la mejora del nuevo tratamiento, que tiene una respuesta promedio 6.5 unidades mayor que la del tratamiento control. Sin embargo, en un estudio más a fondo, se observa que su varianza es mayor que la del tratamiento control. El hecho que se incremente la varianza cuando lo hace la media poblacional es una tendencia usual de los fenómenos industriales, así como en la naturaleza. Para describir esta situación realista, la Figura 3.5 muestra la gráfica de una tercera función de densidad normal, con media $\mu_2 = 16.5$ y varianza $\sigma^2 = 3.5^2 = 12.25$. Con lo cual, se mejora la respuesta media con el efecto secundario de una mayor incertidumbre.

Una importante, y evidente, propiedad de la distribución normal es la simetría respecto del promedio poblacional μ. Una variable aleatoria X o su distribución de probabilidad es *simétrica* respecto de cierta constante $\mu \in \mathbf{R}$, si las variables aleatorias $X - \mu$ y $\mu - X$ tienen la misma distribución:

$$P(X - \mu \leq x) = P(\mu - X \leq x); \quad \text{para todo} \quad x \in \mathbf{R}.$$

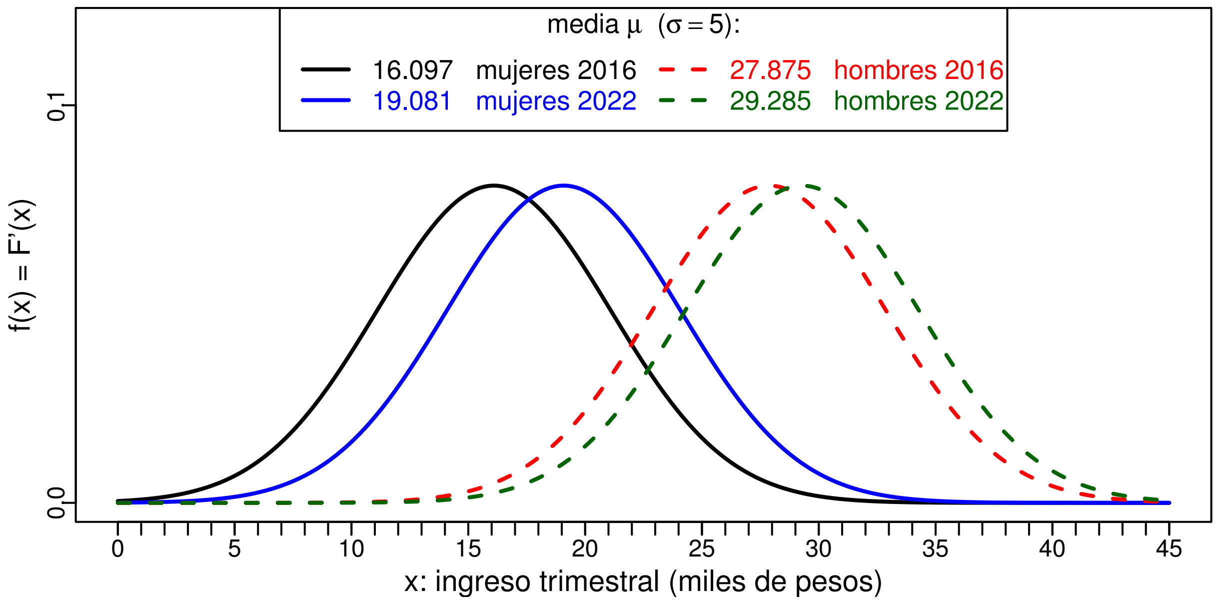


Figura 3.4. Gráfica de cuatro funciones de densidad normal, de medias poblacionales en Tabla 3.1.

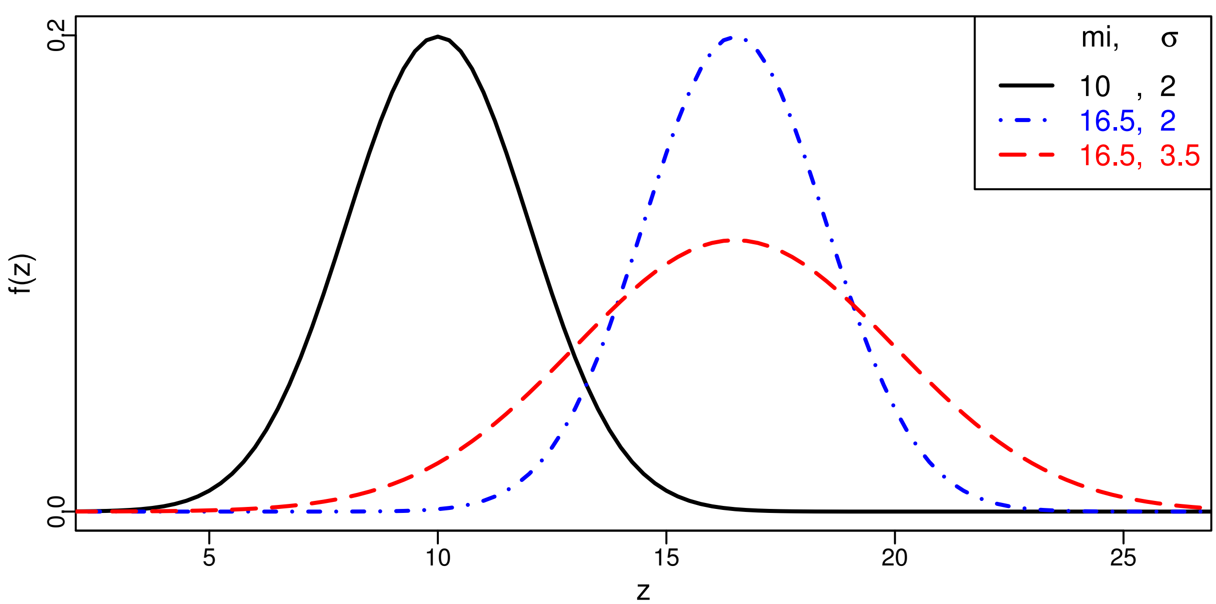


Figura 3.5. Gráfica de tres funciones de densidad normal, dos de ellas con igual varianza ($\sigma_1^2 = \sigma_2^2 = 2^2 = 4$) y dos con igual media poblacional ($\mu_2 = \mu_3 = 16.5$).

Las siguientes afirmación equivalen a que la variable aleatoria X sea simétrica respecto de μ

$$\begin{aligned}
&P(X-\mu \leq x) = P(\mu - X \leq x),\\
&P(X \leq \mu + x) = P(X \geq \mu - x),\\
&P(X < \mu + x) = P(X > \mu - x),\\
&P(X < \mu - x) = P(X > \mu - x),\\
&P(X \leq \mu + x) = P(X \geq \mu + x); \quad \text{para todo} \quad x \in \mathbf{R}.
\end{aligned}$$

En tal caso, el punto μ es único y es mediana poblacional. Además, este valor es la media de la población siempre que exista. Para el caso de una variable aleatoria continua o discreta, con función de densidad $f(x)$, la simetría de su población equivale a que $f(x)$ sea una función *simétrica* respecto de μ:

$$f(\mu - x) = f(\mu + x); \quad \text{para todo} \quad x \in \mathbf{R}.$$

Véase el Ejercicio 3.14. Por ejemplo, una población de distribución normal es simétrica respecto de su media μ; véase Ejercicio 3.17. En particular, para el caso de la distribución normal estándar se tiene

$$\phi(-z) = \frac{1}{\sqrt{2\pi}} e^{-(-z)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \phi(z); \quad \text{para todo} \quad z \in \mathbf{R}. \tag{3.33}$$

Por lo tanto

$$Z \sim \mathrm{N}(0,1) \quad \text{si y sólo si} \quad -Z \sim \mathrm{N}(0,1). \tag{3.34}$$

De hecho, por (3.33), se tiene

$$\begin{aligned}
P(-Z \leq z) = P(Z \geq -z) &= \int_{-z}^{\infty} \phi(t)\mathrm{d}t\\
&= \int_{z}^{-\infty} \phi(-u)(-\mathrm{d}u) = -\int_{z}^{-\infty} \phi(u)\mathrm{d}u\\
&= \int_{-\infty}^{z} \phi(u)\mathrm{d}u = \Phi(z)\\
&= P(Z \leq z); \quad z \in \mathbf{R}.
\end{aligned}$$

Aquí se usó el cambio de variable $u = -t$. Como consecuencia, se obtienen las siguientes dos identidades equivalentes

$$\Phi(-z) = 1 - \Phi(z), \tag{3.35}$$

$$\Phi(z) = 1 - \Phi(-z); \quad z \in \mathbf{R}. \tag{3.36}$$

Esta identidad es una propiedad de cualquier función de distribución de una variable aleatoria continua simétrica respecto de $z = 0$. Es suficiente verificar la primera identidad:

$$\Phi(-z) = P(Z \leq -z) = P(-Z \leq -z) = P(Z \geq z) = 1 - \Phi(z); \quad z \in \mathbf{R}.$$

En general, las identidades (3.35) y (3.36) se satisfacen para cualquier función de distribución continua $F(x)$ de una variable aleatoria continua simétrica alrededor de cero; véase el Ejercicio

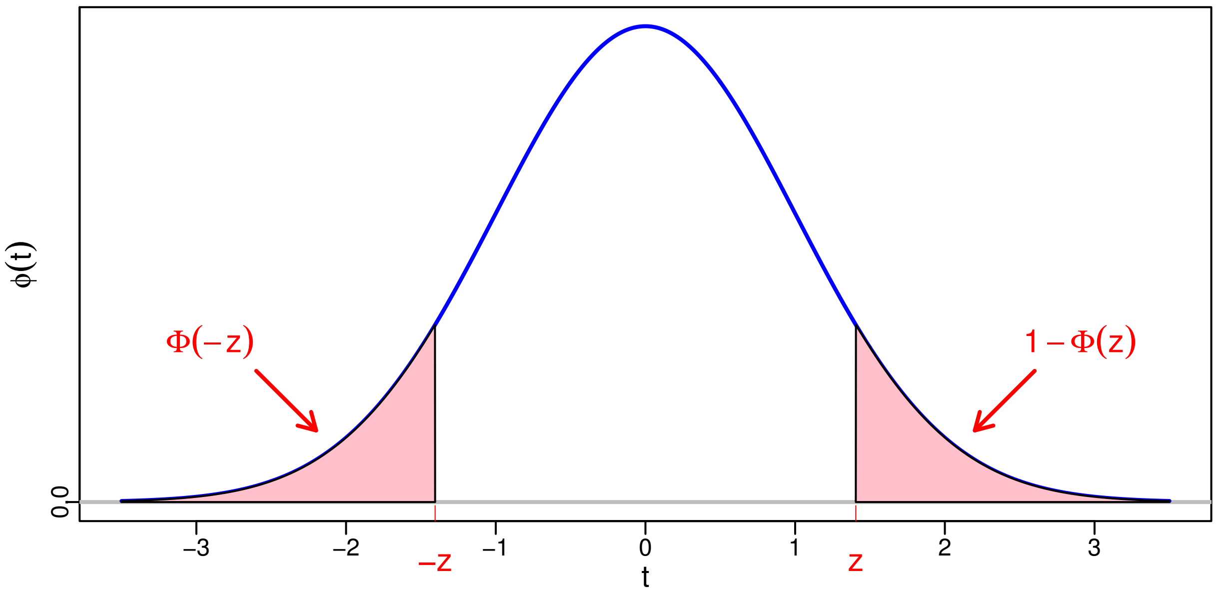


Figura 3.6. Representación gráfica de la identidad (3.35).

3.15. En la Figura 3.6 se aprecia una versión gráfica de la primera identidad. El área bajo la curva de la función de densidad normal estándar $\phi(t)$, en el intervalo $(-\infty, -z)$, es la misma que el área correspondiente del intervalo (z, ∞); con $z > 0$.

Por otro lado, la simetría de la función de densidad normal estándar implica

$$P(-z < Z < z) = 2\Phi(z) - 1; \quad z \geq 0, \tag{3.37}$$

De hecho, para $z \geq 0$, resulta

$$\begin{aligned} P(-z < Z < z) &= P(Z < z) - P(Z \leq -z) \\ &= \Phi(z) - \Phi(-z) \\ &= \Phi(z) - [1 - \Phi(z)] \\ &= 2\Phi(z) - 1. \end{aligned}$$

La familia de distribuciones normales es cerrada bajo transformación de localización y escala: si $a \in \mathbf{R}$ y $b > 0$, entonces

$$X \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Y = a + bX \sim \mathrm{N}(a + b\mu, b^2\sigma^2). \tag{3.38}$$

Este resultado aplica incluso para $b < 0$, véase el Ejercicio 3.20. Recuerde también

$$\mathrm{E}\,Y = \mathrm{E}(a + bX) = a + b\,\mathrm{E}\,X \quad \text{y} \quad \mathrm{V}\,Y = \mathrm{V}(a + bX) = b^2\,\mathrm{V}\,X.$$

En particular, para $\mu = a = 0$ y $\sigma = -b = 1$, se obtiene la relación (3.34).

Algunos ejemplos de cálculo de probabilidades son los siguientes. Se puede apoyar en las tablas de probabilidad normal en [9, Apédice A], o bien con la función *qnorm()* del software estadístico R [14].

$$\begin{aligned}
P(Z < -1) &= \Phi(-1) = 0.1587 = 1 - 0.8413 = 1 - \Phi(1) = P(Z > 1),\\
P(Z < -2) &= \Phi(-2) = 0.0228 = 1 - 0.9772 = 1 - \Phi(2) = P(Z > 2),\\
P(Z < -3) &= \Phi(-3) = 0.0014 = 1 - 0.9986 = 1 - \Phi(3) = P(Z > 3),\\
P(Z \leq -3.99) &= \Phi(-3.99) = 0.000033 = 1 - 0.999967 = 1 - \Phi(3.99) = P(Z > 3.99),\\
P(Z \leq 1.73) &= \Phi(1.73) = 0.9582 = 1 - 0.0418 = 1 - \Phi(-1.73),\\
P(Z \leq 1.74) &= \Phi(1.74) = 0.9591 = 1 - 0.0409 = 1 - \Phi(-1.74).
\end{aligned}$$

Cuando el valor posible z tiene tres o más decimales, se sustituye por cualquiera de los siguientes valores posibles de la tabla de probabilidad normal: (a) el más cercano, (b) el más cercano a su izquierda, (c) el más cercano a su derecha, (d) el promedio de los dos más cercanos. Por ejemplo

$$\begin{aligned}
P(Z \leq 1.736) &\approx P(Z \leq 1.74) = \Phi(1.74) = 0.9591,\\
P(Z \leq 1.736) &\approx \frac{1}{2}[\Phi(1.73) + \Phi(1.74)] = \frac{1}{2}(0.9582 + 0.9591) = 0.9587.
\end{aligned}$$

Al aplicar (3.37), se tiene

$$\begin{aligned}
P(-1 < Z < 1) &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 2(0.8413) - 1 = 0.6826,\\
P(-2 < Z < 2) &= 2\Phi(2) - 1 = 2(0.9772) - 1 = 0.9544,\\
P(-3 < Z < 3) &= 2\Phi(3) - 1 = 2(0.9987) - 1 = 0.9974,\\
P(-4 < Z < 4) &= 2\Phi(4) - 1 = 2(0.99997) - 1 = 0.99994.
\end{aligned}$$

Por lo tanto, en una población normal, el 68.26 % de sus individuos se encontrará entre -1 y 1. Así mismo, entre -2 y 2 se localiza el 95.44 % de los individuos. El 99.74 % de la población pertenecerá al intervalo $(-3, 3)$. Como consecuencia, un individuo de la población con valor $Z = 3.27$, es atípico o inusual.

Dada una probabilidad $\alpha \in (0, 1)$, hay un único punto $z \in \mathbf{R}$, tal que

$$\Phi(z) = P(Z \leq z) = \alpha,$$

Dicho punto, que se denota como z_α, es el *cuantil α*, *α-cuantil* o *α-cuantil inferior* de la distribución normal estándar. Así

$$z_\alpha = \Phi^{-1}(\alpha); \quad \text{para} \quad 0 < \alpha < 1.$$

De hecho, la función $\Phi(z)$ en (3.25) es estrictamente creciente e invertible, pues es la integral de una función de densidad positiva. El dominio e imagen de la función inversa $\Phi^{-1}(\alpha)$ es la imagen y dominio de $\Phi(z)$; respectivamente:

$$\Phi^{-1} : (0, 1) \to (-\infty, \infty).$$

En general, si X es una variable aleatoria continua, con conjunto de valores posibles (a, b); para ciertos $-\infty \leq a < b \leq \infty$, y función de distribución $F(x)$, estrictamente creciente en (a, b), entonces $F(x)$ es invertible:

$$F^{-1} : (0, 1) \to (a, b).$$

Este es el caso donde $F(x)$ es diferenciable en (a,b), con función de densidad positiva en el mismo intervalo; casi en todas partes. La función inversa $F^{-1}(\alpha)$ se denota también como x_α:

$$x_\alpha = F^{-1}(\alpha) \in (a,b), \quad \text{si y sólo si} \quad F(x_\alpha) = P(X \leq x_\alpha) = \alpha \in (0,1). \tag{3.39}$$

Al valor posible x_α se le dice *cuantil* α, α*-cuantil* o α*-cuantil inferior* de la distribución de la variable aleatoria X. El α-cuantil x_α se asocia a la probabilidad de la cola izquierda. De manera similar, se define el α*-cuantil superior*, que corresponde a la probabilidad de la cola derecha

$$1 - F(x_\alpha) = P(X > x_\alpha) = \alpha; \quad \text{para} \quad 0 < \alpha < 1.$$

Sin embargo, hay que tener cuidado en su uso, pues se denota también como x_α. Para el caso de la distribución normal estándar, z_α denota usualmente el α-cuantil superior; para $\alpha \in (0,1)$ dado. Por ejemplo, $z_{0.05}$ es el 0.05-cuantil superior de la distribución normal estándar:

$$\begin{aligned} P(Z > z_{0.05}) &= 0.05, \\ P(Z \leq z_{0.05}) &= 0.95, \\ \Phi(z_{0.05}) &= 0.95, \\ z_{0.05} &= \frac{1.64 + 1.65}{2} = 1.645. \end{aligned}$$

Su comprobación es inmediata

$$\Phi(1.645) = \frac{0.9495 + 0.9505}{2} = 0.95.$$

Encuentre el intervalo $(-z, z)$, de cobertura 95 %, de la distribución normal estándar. Por (3.37), hay que resolver

$$\begin{aligned} P(-z < Z < z) &= 0.95, \\ 2\Phi(z) - 1 &= 0.95, \\ \Phi(z) &= \frac{1 + 0.95}{2} = 0.975, \\ z &= \Phi^{-1}(0.975) = 1.96. \end{aligned}$$

Por lo tanto, un intervalo de predicción del 95 % de la distribución normal estándar es

$$-1.96 < Z < 1.96.$$

Este intervalo contiene al 95 % de los individuos de una población normal estándar; al rededor de su centro cero. Por la simetría de la función de densidad, el 2.5 % de los individuos se encontrará en su cola inferior:

$$P(Z \leq -1.96) = 0.025.$$

Mientras que un porcentaje igual de individuos estará en la cola superior:

$$P(Z > 1.96) = 0.025.$$

α	z_α (cuantil superior)	$\Phi(z_\alpha)$	$1-\Phi(z_\alpha)$
0.01	2.3263	0.99	0.01
0.025	1.96	0.975	0.025
0.05	1.6449	0.95	0.05
0.1	1.2816	0.9	0.1
0.5	0	0.5	0.5

Tabla 3.2. Algunos cuantiles superiores de la distribución normal estándar.

Así, el 0.025-cuantil superior es

$$z_{0.025} = 1.96.$$

Un intervalo de predicción de $100 \times (1-\alpha)\,\%$ es

$$-z_{\alpha/2} < Z < z_{\alpha/2}; \quad \text{para} \quad 0 < \alpha < 1,$$

donde $z_{\alpha/2}$ es el $\alpha/2$-cuantil superior de la distribución normal estándar. En la Tabla 3.2 se muestran algunos de los cuantiles superiores de la distribución normal estándar.

Ejemplo 3.5.1. Considere un activo con riesgo cuya ganancia (lotería) X por cada peso invertido tiene una distribución normal de media $\mu = 0.043$ pesos y desviación estándar $\sigma = 0.02$ pesos. Obtenga la probabilidad de los siguientes eventos

(a) Pérdida $\{X < 0\}$,

(b) Ganancia mayor que 8.92 centavos,

(c) Ganancia entre 3.1 y 8.55 centavos

Obtenga también los siguientes intervalos de predicción de 90 % del rendimiento

(d) Al rededor del rendimiento medio

(e) Cola derecha

(f) Cola izquierda

Solución. a) Al cambiar la escala de la variable de interés en centavos, se tiene

$$\mu = \operatorname{E} X = 4.3, \quad \sigma = \text{d. e.}(X) = 2 \quad \text{y} \quad \sigma^2 = V X = 4,$$

donde X denota la ganancia en centavos del activo con riesgo por cada peso invertido. El modelo de distribución normal es

$$X \sim \operatorname{N}(\mu = 4.3, \sigma^2 = 4).$$

Este modelo equivale a su versión estandarizada

$$Z = \frac{X-\mu}{\sigma} = \frac{X-4.3}{2} \sim \operatorname{N}(0,1).$$

Así

$$\begin{aligned}
P(X<0) &= P\left(\frac{X-\mu}{\sigma} < \frac{0-\mu}{\sigma}\right) \\
&= P\left(\frac{X-4.3}{2} < \frac{0-4.3}{2}\right) \\
&= P(Z<-2.15) = \Phi(-2.15) \\
&= 0.0158.
\end{aligned}$$

b) Se cumplen las siguientes igualdades del evento de interés

$$\begin{aligned}
\{3.1<X<8.55\} &= \left\{\frac{3.1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{8.55-\mu}{\sigma}\right\} \\
&= \left\{\frac{3.1-4.3}{2} < Z < \frac{8.55-4.3}{2}\right\} \\
&= \{-0.6<Z<2.125\}.
\end{aligned}$$

Por lo tanto

$$\begin{aligned}
P(3.1<X<8.55) &= P(-0.6<Z<2.125) = \Phi(2.125) - \Phi(-0.6) \\
&= \frac{0.983+0.9834}{2} - 0.2743 = 0.9832 - 0.2743 \\
&= 0.7089.
\end{aligned}$$

c) De manera similar

$$\begin{aligned}
P(X>8.92) &= P\left(\frac{X-\mu}{\sigma} > \frac{8.92-\mu}{\sigma}\right) \\
&= P\left(\frac{X-4.3}{2} < \frac{8.92-4.3}{2}\right) \\
&= P(Z>2.31) \\
&= 1-P(Z\leq 2.31) \\
&= 1-\Phi(2.31) \\
&= 1-0.9896 \\
&= 0.0104.
\end{aligned}$$

d) Aquí se requieren los valores $x_1, x_2 \in \mathbf{R}$ tal que

$$P(x_1<X<x_2) = 0.9.$$

A las colas derecha e izquierda les corresponde una probabilidad del 0.1. Por lo que a cada una de ellas tiene probabilidad 0.05. De nuevo, se estandariza el evento de interés:

$$\begin{aligned}\{x_1 < X < x_2\} &= \left\{\frac{x_1-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}\right\}\\ &= \left\{\frac{x_1-4.3}{2} < Z < \frac{x_2-4.3}{2}\right\}\\ &= \left\{-z_{\alpha/2} < Z < z_{\alpha/2}\right\}\\ &= \left\{-z_{0.05} < Z < z_{0.05}\right\},\end{aligned}$$

donde

$$\begin{aligned}z_{\alpha/2} &= z_{0.05}\\ &= 0.05\text{-cuantil superior}\\ &= 1.6449.\end{aligned}$$

Entonces

$$\begin{aligned}\frac{x_1-\mu}{\sigma} &= -1.6449, & \frac{x_2-\mu}{\sigma} &= 1.6449,\\ x_1 &= \mu - \sigma z_{\alpha/2}, & x_2 &= \mu + \sigma z_{\alpha/2},\\ x_1 &= 4.3 - 2(1.6449), & x_2 &= 4.3 + 2(1.6449),\\ x_1 &= 4.3 - 3.2898, & x_2 &= 4.3 + 3.2898,\\ x_1 &= 1.0102, & x_2 &= 7.5898.\end{aligned}$$

Por lo tanto, un intervalo de 90 % de predicción del rendimiento al rededor de la media es $1.01 < X < 7.59$ centavos.

$$1.0102 < X < 7.5898 \text{ centavos.}$$

e) El intervalo de predicción requerido tiene la forma $(-\infty, x_2)$, donde el cuantil x_2 se asocia al 0.1-cuantil superior de la distribución normal estándar:

$$\begin{aligned}\frac{x_2-\mu}{\sigma} &= z_\alpha,\\ x_2 &= \mu + \sigma z_\alpha,\\ x_2 &= 4.3 + 2z_{0.1},\\ x_2 &= 4.3 + 2(1.2816),\\ x_2 &= 4.3 + 2.5632,\\ x_2 &= 6.5632.\end{aligned}$$

Por lo tanto, el 90 % de los rendimientos factibles se encontrarán en el intervalo $X < 6.5632$ centavos.

f) Por último, el intervalo de predicción requerido tiene la forma (x_1, ∞), donde el cuantil x_1

se asocia al 0.1-cuantil inferior de la distribución normal estándar:

$$
\begin{aligned}
\frac{x_1 - \mu}{\sigma} &= -z_\alpha, \\
x_1 &= \mu - \sigma z_\alpha, \\
x_1 &= 4.3 - 2z_{0.1}, \\
x_1 &= 4.3 - 2(1.2816), \\
x_1 &= 4.3 - 2.5632, \\
x_1 &= 1.7368.
\end{aligned}
$$

Por lo tanto, el 90 % de los rendimientos factibles se encontrarán en el intervalo $X > 1.7368$ centavos.

Ahora se deducirá la función generadora de momentos de la distribución normal. Si X es una variable aleatoria de distribución normal de media μ y varianza σ^2, su función generadora de momentos es

$$M(t) = \mathrm{E}\, e^{tX} = e^{\mu t + \sigma^2 t^2/2}; \quad \text{para} \quad t \in \mathbf{R}. \tag{3.40}$$

Se verifica primero el caso normal estándar. Si $Z \sim \mathrm{N}(0, 1)$, entonces

$$
\begin{aligned}
M_Z(t) = \mathrm{E}\, e^{tZ} &= \int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz - z^2/2} \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(t^2 - 2tz + z^2)/2 + t^2/2} \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t^2/2} e^{-(z-t)^2/2} \mathrm{d}z = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} \mathrm{d}z \\
&= e^{t^2/2}.
\end{aligned}
$$

El último integrando corresponde a una función de densidad normal de media t y varianza uno. Así que su integral es uno. Ahora, para el caso general, considere la relación (3.38):

$$X \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad X = \mu + \sigma Z, \quad \text{con} \quad Z \sim \mathrm{N}(0, 1).$$

Así que

$$
\begin{aligned}
M(t) = \mathrm{E}\, e^{tX} &= \mathrm{E}\, e^{t(\mu + \sigma Z)} \\
&= \mathrm{E}[e^{t\mu} e^{t\sigma Z}] = e^{t\mu}\, \mathrm{E}[e^{(t\sigma)Z}] \\
&= e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} \\
&= e^{\mu t + \sigma^2 t^2/2}.
\end{aligned}
$$

La fórmula (3.40) se deduce también por la primera propiedad de las funciones generadoras de momentos del Teorema 2.6.8:

$$M(t) = \mathrm{E}\, e^{tX} = \mathrm{E}\, e^{t(\mu + \sigma Z)} = M_{\mu + \sigma Z}(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t + \sigma^2 t^2/2}; \quad t \in \mathbf{R}.$$

Una aplicación de (3.40) es la obtención de los momentos tanto de la distribución normal como log normal. Una variable aleatoria Y tiene distribución *log normal* si $X = \log Y$ tiene distribución normal:

$$Y = e^X \sim \mathrm{logN}(\mu, \sigma^2) \quad \text{si y sólo si} \quad X \sim \mathrm{N}(\mu, \sigma^2).$$

Con lo que

$$\mathrm{E}\, Y^k = e^{\mu k + \sigma^2 k^2/2}; \quad \text{para} \quad k = 0, 1, \ldots. \tag{3.41}$$

De hecho

$$\mathrm{E}\, Y^k = \mathrm{E}[(e^X)^k] = \mathrm{E}[e^{kX}] = M_X(k) = e^{\mu k + \sigma^2 k^2/2}.$$

Con esta resultado se obtienen los momentos de las distribuciones normal y log normal.

$$\begin{aligned}
\mathrm{E}\, Y &= e^{\mu(1) + \sigma^2 (1)^2/2} = e^{\mu + \sigma^2/2}, \\
\mathrm{E}\, Y^2 &= e^{\mu(2) + \sigma^2 (2)^2/2} = e^{2\mu + 2\sigma^2}, \\
\mathrm{E}\, Y^3 &= e^{\mu(3) + \sigma^2 (3)^2/2} = e^{3\mu + (9/2)\sigma^2}, \\
\mathrm{E}\, Y^4 &= e^{\mu(4) + \sigma^2 (4)^2/2} = e^{4\mu + 8\sigma^2}.
\end{aligned}$$

Así

$$\begin{aligned}
\mathrm{V}\, Y &= \mathrm{E}\, Y^2 - (\mathrm{E}\, Y)^2 = e^{2\mu + 2\sigma^2} - e^{2(\mu + \sigma^2/2)} \\
&= e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} = e^{2\mu + \sigma^2}[e^{\sigma^2} - 1],
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(Y - \mathrm{E}\, Y)^3 &= \mathrm{E}\, Y^3 - 3\, \mathrm{E}\, Y\, \mathrm{E}\, Y^2 + 2(\mathrm{E}\, Y)^3 \\
&= e^{3\mu + (9/2)\sigma^2} - 3e^{\mu + \sigma^2/2} e^{2\mu + 2\sigma^2} + 2e^{3(\mu + \sigma^2/2)} \\
&= e^{3\mu + (3/2)\sigma^2}[e^{3\sigma^2} - 3e^{\sigma^2} + 2] \\
&= e^{3\mu + (3/2)\sigma^2}[e^{\sigma^2} + 2][e^{\sigma^2} - 1]^2,
\end{aligned}$$

y

$$\begin{aligned}
\mathrm{E}(Y - \mathrm{E}\, Y)^4 &= \mathrm{E}\, Y^4 - 4\, \mathrm{E}\, Y\, \mathrm{E}\, Y^3 + 6(\mathrm{E}\, Y)^2\, \mathrm{E}\, Y^2 - 3(\mathrm{E}\, Y)^4 \\
&= e^{4\mu + 8\sigma^2} - 4e^{\mu + \sigma^2/2} e^{3\mu + (9/2)\sigma^2} + 6e^{2(\mu + \sigma^2/2)} e^{2\mu + 2\sigma^2} - 3e^{4(\mu + \sigma^2/2)} \\
&= e^{4\mu + 2\sigma^2}[e^{6\sigma^2} - 4e^{3\sigma^2} + 6e^{\sigma^2} - 3] \\
&= e^{4\mu + 2\sigma^2}[e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3](e^{\sigma^2} - 1)^2.
\end{aligned}$$

Por lo que

$$\begin{aligned}
\mathrm{sesgo}(Y) &= \frac{\mathrm{E}(Y - \mathrm{E}\, Y)^3}{(\mathrm{V}\, Y)^{3/2}} = \frac{e^{3\mu + (3/2)\sigma^2}(e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^2}{[e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)]^{3/2}} \\
&= \frac{e^{3\mu + (3/2)\sigma^2}}{e^{3\mu + (3/2)\sigma^2}}(e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^{2 - 3/2} = (e^{\sigma^2} + 2)(e^{\sigma^2} - 1)^{1/2},
\end{aligned}$$

$$
\begin{aligned}
\operatorname{curtosis}(Y) &= \frac{\mathrm{E}(Y-\mathrm{E}\,Y)^4}{(\mathrm{V}\,Y)^{4/2}} \\
&= \frac{e^{4\mu+2\sigma^2}(e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3)(e^{\sigma^2}-1)^2}{[e^{2\mu+\sigma^2}(e^{\sigma^2}-1)]^2} \\
&= \frac{e^{4\mu+2\sigma^2}}{e^{4\mu+2\sigma^2}}(e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3) \\
&= e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3.
\end{aligned}
$$

y

$$
\begin{aligned}
\operatorname{curtosis.e.}(Y) &= e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3-3 \\
&= e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6 \\
&= (e^{3\sigma^2}+3e^{2\sigma^2}+6e^{\sigma^2}+6)(e^{\sigma^2}-1).
\end{aligned}
$$

Observe que tanto la media como la varianza de la variable aleatoria log normal Y dependen de los parámetros μ y σ^2. Dichos indicadores poblacionales son cada vez mayores conforme μ o σ^2 crecen. Sin embargo, el sesgo y la curtosis dependen sólo de σ^2; con una relación directa. Esto último se explica porque la variable aleatoria Y se representa como

$$Y = e^X \overset{d}{=} e^{\mu+\sigma Z} = e^{\mu}e^{\sigma Z}, \quad \text{con} \quad Z \sim \mathrm{N}(0,1).$$

Así

$$Y \overset{d}{=} \theta e^{\sigma Z}, \quad \text{con} \quad \theta = e^{\mu} > 0.$$

Como Y es una transformación de escala de $e^{\sigma Z}$, entonces Y y $e^{\sigma Z}$ comparten el sesgo y la curtosis. Por último, cabe destacar que $\theta = e^{\mu} > 0$ es un parámetro de escala de la distribución log normal. Así que el parámetro $\sigma > 0$ modula la forma, la cual es desde casi simétrica, con curtosis cercana a 3 cuando $\sigma \approx 0$, hasta lo que se desee de asimetría y cola derecha pesada, para $\sigma \to \infty$.

Los primeros dos momentos de la variable aleatoria X se obtienen al derivar en $t = 0$ su función generadora de momentos.

$$
\begin{aligned}
M(t) &= \mathrm{E}\,e^{tX} = e^{\mu t+\sigma^2t^2/2}, \\
M'(t) &= (\mu+\sigma^2 t)e^{\mu t+\sigma^2t^2/2} = (\mu+\sigma^2 t)M(t), \qquad (3.42)\\
M''(t) &= \sigma^2 M(t) + (\mu+\sigma^2 t)M'(t) = [\sigma^2+(\mu+\sigma^2 t)^2]M(t). \qquad (3.43)
\end{aligned}
$$

Así

$$
\begin{aligned}
M(0) &= e^{\mu(0)+\sigma^2(0)^2/2} = 1, \\
\mathrm{E}\,X &= M'(0) = (\mu+\sigma^2(0))M(0) = \mu, \\
\mathrm{E}\,X^2 &= M''(0) = [\sigma^2+(\mu+\sigma^2(0))^2]M(0) = \mu^2+\sigma^2, \\
\mathrm{V}\,X &= \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = (\mu^2+\sigma^2)-\mu^2 = \sigma^2.
\end{aligned}
$$

Por su simetría, el sesgo de un población normal es cero. Además, su curtosis es 3; véase el Ejercicio 3.21.

Recuerde que el coeficiente de sesgo de una variable aleatoria, que se obtiene con el tercer momento central, aporta información sobre la asimetría de una población. Así mismo, con el cuarto momento central se obtiene el coeficiente de curtosis, el cual mide la pesadez de las colas. En ocasiones, sin que implique una regla, una distribución con colas pesadas induce una forma picuda en el centro. Con lo cual, el coeficiente de curtosis identifica indirectamente formas picudas o llanas de la distribución. Por definición, tanto el coeficiente de sesgo como de curtosis no dependen de la escala o unidad de medida de la variable aleatoria de interés. En particular, dichos coeficientes son iguales a su contraparte de la variable aleatoria estandarizada:

$$\begin{aligned} \text{sesgo}(X) &= \text{sesgo}\left(\frac{X-\mu}{\sigma}\right) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3, \\ \text{curtosis}(X) &= \text{curtosis}\left(\frac{X-\mu}{\sigma}\right) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^4, \end{aligned}$$

donde $\mu = \mathrm{E}\,X$ y $\sigma^2 = \mathrm{V}\,X$. Por último, la *curtosis excedente* es

$$\text{curtosis. e.}(X) = \text{curtosis}(X) - 3.$$

En particular una población normal tiene curtosis excedente cero.

Distribución uniforme continua. Una variable aleatoria X tiene distribución *uniforme continua* en el intervalo (a, b), con $-\infty < a < b < \infty$, si su función de densidad es

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{o.c.} \end{cases}$$

o bien

$$f(x) = \frac{1}{b-a}; \quad \text{para} \quad a < x < b.$$

Se denota como

$$X \sim \mathrm{U}(a, b).$$

El conjunto de valores posibles de X es el intervalo $\mathcal{X} = (a, b)$. La característica básica de una población uniforme es de que la probabilidad de encontrar un individuo en una región B, subconjunto del intervalo (a, b), es proporcional a su longitud:

$$P(X \in B) = c \times \text{longitud}(B),$$

donde $c > 0$ es la constante de proporcionalidad. Como

$$P(X \in (a, b)) = c \times \text{longitud}(a, b) = c(b-a) = 1,$$

entonces

$$c = \frac{1}{b-a}.$$

Así
$$P(X \in B) = \frac{\text{longitud}(B)}{b-a}, \quad B \subset (a,b).$$
Por ejemplo, si $B = ((a+b)/2, b)$, entonces
$$\begin{aligned}
P(X \in B) &= \int_B f(t)\mathrm{d}t = \int_{(a+b)/2}^{b} \frac{1}{b-a}\mathrm{d}t \\
&= \frac{1}{b-a}\int_{(a+b)/2}^{b} \mathrm{d}t = \frac{1}{b-a}[t]_{(a+b)/2}^{b} \\
&= \frac{1}{b-a}\left(b - \frac{a+b}{2}\right) = \frac{2b-(a+b)}{2(b-a)} \\
&= \frac{b-a}{2(b-a)} = \frac{1}{2}.
\end{aligned}$$
La función de distribución es
$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$
o bien, en forma simplificada
$$F(x) = \frac{x-a}{b-a}, \quad a \leq x \leq b.$$
De hecho, para $a \leq x \leq b$, se tiene
$$\begin{aligned}
F(x) = P(X \leq x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{a} f(t)\mathrm{d}t + \int_{a}^{x} f(t)\mathrm{d}t = \int_{-\infty}^{a} 0\mathrm{d}t + \int_{a}^{x}\frac{1}{b-a}\mathrm{d}t \\
&= \frac{1}{b-a}[t]_a^x = \frac{x-a}{b-a}.
\end{aligned}$$
La media y varianza de esta población es
$$\mu = \mathrm{E}\,X = \frac{a+b}{2} \quad \text{y} \quad \sigma^2 = \mathrm{V}\,X = \frac{(b-a)^2}{12}.$$
Por ejemplo, una rueda de radio uno tiene una circunferencia de longitud 2π. Así $a = 0$, $b = 2\pi$ y
$$\begin{aligned}
\mu = \mathrm{E}\,X &= \frac{0+2\pi}{2} = \pi, \\
\sigma^2 = \mathrm{V}\,X &= \frac{(2\pi-0)^2}{12} = \frac{\pi^2}{3}.
\end{aligned}$$

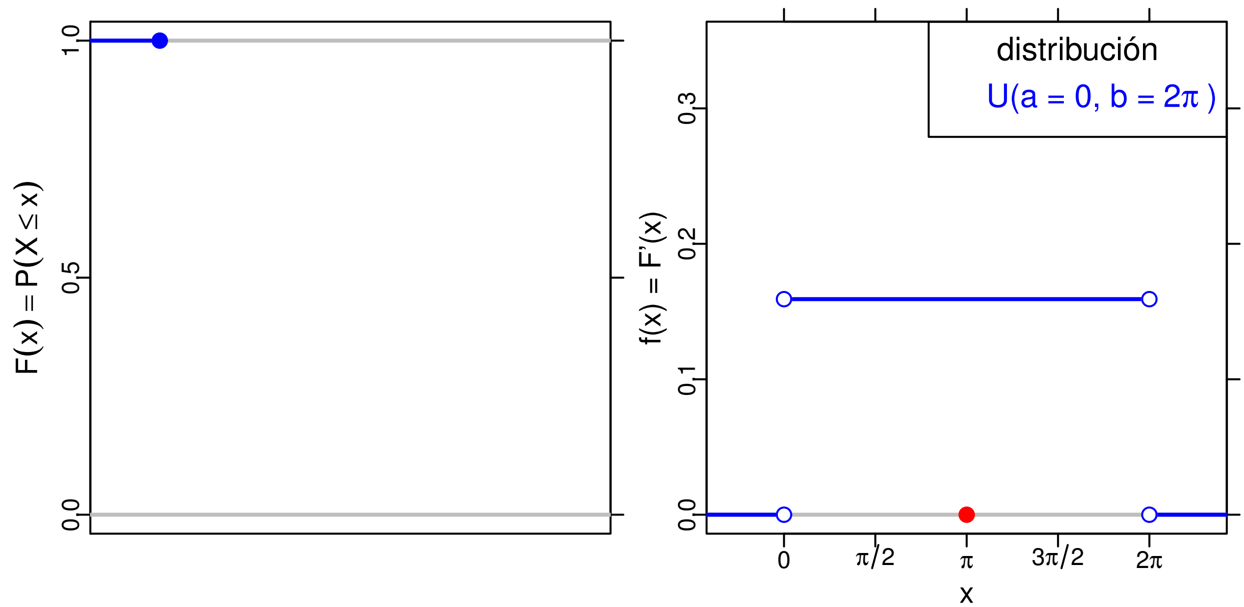


Figura 3.7. [arriba] Función de densidad uniforme en el intervalo $(0, 2\pi)$. [abajo] Correspondiente función de distribución.

Las gráficas de la función de densidad y función de distribución de este ejemplo, se muestran en la Figura 3.7.

Los momentos no centrales de la distribución uniforme son

$$EX^k = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}; \quad \text{para} \quad k = 0, 1, 2, \dots \tag{3.44}$$

De hecho

$$\mathrm{E}\, X^k = \int_{-\infty}^{\infty} x^k f(x) \mathrm{d}x = \int_a^b \frac{x^k}{b-a} \mathrm{d}x = \frac{[x^{k+1}]_a^b}{(k+1)(b-a)} = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}.$$

El sesgo de una población uniforme continua es cero, pues es simétrica al rededor de la media $\mu = (a+b)/2$. La curtosis debe menor que la de una población normal: $0 < \text{curtosis}(X) < 3$; pues están recortadas ambas colas de la población, derecha e izquierda. Sin pérdida de generalidad, se verificarán estas afirmaciones para el caso $X \sim \mathrm{U}(0,1)$. Por (3.44), se tiene

$$\mathrm{E}\, X^k = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)} = \frac{1}{k+1}; \quad \text{para} \quad k = 0, 1, 2, \dots$$

Así

$$\mu = \mathrm{E}\, X = \frac{1}{2},$$

$$\mathrm{E}\, X^2 = \frac{1}{3},$$

$$\sigma^2 = \mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12},$$

$$
\begin{aligned}
\operatorname{E} X^3 &= \frac{1}{4},\\
\operatorname{E}(X-\mu)^3 &= \operatorname{E} X^3 - 3\mu \operatorname{E} X^2 + 2\mu^3\\
&= \frac{1}{4} - \frac{3}{2}\left(\frac{1}{3}\right) + 2\left(\frac{1}{2}\right)^3 = 0,\\
\operatorname{sesgo}(X) &= \operatorname{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{\operatorname{E}(X-\mu)^3}{\sigma^3} = 0
\end{aligned}
$$

y

$$
\begin{aligned}
\operatorname{E} X^4 &= \frac{1}{5},\\
\operatorname{E}(X-\mu)^4 &= \operatorname{E} X^4 - 4\mu \operatorname{E} X^3 + 6\mu^2 \operatorname{E} X^2 - 3\mu^4\\
&= \frac{1}{5} - \frac{4}{2}\left(\frac{1}{4}\right) + 6\left(\frac{1}{2}\right)^2 \frac{1}{3} - 3\left(\frac{1}{2}\right)^4 = \frac{1}{80},\\
\operatorname{curtosis}(X) &= \operatorname{E}\left(\frac{X-\mu}{\sigma}\right)^4 = \frac{\operatorname{E}(X-\mu)^4}{(\sigma^2)^2} = \frac{1/80}{(1/12)^2} = 1.8 < 3,\\
\operatorname{curtosis.e.}(X) &= \operatorname{curtosis}(X) - 3 = 1.8 - 3 = -1.2 < 0.
\end{aligned}
$$

La fórmula de momentos (3.44) se verifica también con las derivadas en $t=0$ de la función generadora de momentos. Se aplicarán las primeras cuatro de esas derivadas.

$$
M(t) = \operatorname{E} e^{tX} = \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x = \int_0^1 e^{tx}\mathrm{d}x = \left[\frac{e^{tx}}{t}\right]_0^1 = \frac{e^t - 1}{t}; \quad t \neq 0.
$$

Para $t=0$, la función generadora momentos es uno, valor que corresponde a su límite cuando t tiende a cero:

$$
M(0) = \operatorname{E} X^0 = \lim_{t\to 0} \frac{e^t-1}{t} = \lim_{t\to 0} \frac{e^t}{1} = e^0 = 1.
$$

Aquí se aplicó la regla de L'Hôpital. Luego

$$
\begin{aligned}
M'(t) &= \frac{e^t}{t} - \frac{e^t-1}{t^2} = \frac{e^t - M(t)}{t},\\
M''(t) &= \frac{e^t - M'(t)}{t} - \frac{e^t - M(t)}{t^2}\\
&= \frac{e^t - M'(t)}{t} - \frac{M'(t)}{t} = \frac{e^t - 2M'(t)}{t},\\
M^{(3)}(t) &= \frac{e^t - 2M''(t)}{t} - \frac{e^t - 2M'(t)}{t^2}\\
&= \frac{e^t - 2M''(t)}{t} - \frac{M''(t)}{t} = \frac{e^t - 3M''(t)}{t},\\
M^{(4)}(t) &= \frac{e^t - 3M^{(3)}(t)}{t} - \frac{e^t - 3M''(t)}{t^2}\\
&= \frac{e^t - 3M^{(3)}(t)}{t} - \frac{M^{(3)}(t)}{t} = \frac{e^t - 4M^{(3)}(t)}{t}.
\end{aligned}
$$

Por lo que

$$\begin{aligned} M'(0) &= \lim_{t\to 0} \frac{e^t - M(t)}{t} = \lim_{t\to 0} \frac{e^t - M'(t)}{1} = 1 - M'(0), \\ 2M'(0) &= 1, \\ \mathrm{E}\, X = M'(0) &= \frac{1}{2}, \end{aligned}$$

$$\begin{aligned} M''(0) &= \lim_{t\to 0} \frac{e^t - 2M'(t)}{t} = \lim_{t\to 0} \frac{e^t - 2M''(t)}{1} = 1 - 2M''(0), \\ \mathrm{E}\, X^2 = M''(0) &= \frac{1}{3}, \end{aligned}$$

$$\begin{aligned} M^{(3)}(0) &= \lim_{t\to 0} \frac{e^t - 3M''(t)}{t} = \lim_{t\to 0} \frac{e^t - 3M^{(3)}(t)}{1} = 1 - 3M^{(3)}(0), \\ \mathrm{E}\, X^3 = M^{(3)}(0) &= \frac{1}{4} \end{aligned}$$

y

$$\begin{aligned} M^{(4)}(0) &= \lim_{t\to 0} \frac{e^t - 4M^{(3)}(t)}{t} = \lim_{t\to 0} \frac{e^t - 4M^{(4)}(t)}{1} = 1 - 4M^{(4)}(0), \\ \mathrm{E}\, X^4 = M^{(4)}(0) &= \frac{1}{5}. \end{aligned}$$

Como en el caso de la distribución normal, la familia de distribuciones uniforme es invariante ante transformaciones de localización y escala:

$$X \sim \mathrm{U}(a,b), \quad \text{si y sólo si} \quad Y = c + dX \sim \mathrm{U}(c+ad, c+bd),$$

donde

$$-\infty < a < b < \infty, \quad c \in \mathbf{R}, \quad d > 0.$$

En particular

$$U \sim \mathrm{U}(0,1), \quad \text{si y sólo si} \quad X = a + (b-a)U \sim \mathrm{U}(a,b)$$

y

$$X \sim \mathrm{U}(a,b), \quad \text{si y sólo si} \quad U = \frac{X-a}{b-a} \sim \mathrm{U}(0,1).$$

Distribución gamma. Un modelo de distribución para variables aleatorias positivas es el de la distribución gamma. Esta familia de distribuciones incluye a importantes casos de distribuciones, como exponencial, Erlang o χ-cuadrada.

Una variable aleatoria X tiene *distribución gamma*, de parámetros forma $\alpha > 0$ y escala $\beta > 0$, si su función de densidad es

$$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha \, \Gamma(\alpha)} e^{-x/\beta}; \quad x > 0, \tag{3.45}$$

donde $\Gamma(\alpha)$ es la *función gamma*:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x; \quad \alpha > 0. \tag{3.46}$$

La función gamma satisface las siguientes propiedades

$$\begin{aligned}
\Gamma(1/2) &= \sqrt{\pi}, \\
\Gamma(1) &= 1, \\
\Gamma(2) &= 1, \\
\Gamma(3) &= 2, \\
\Gamma(4) &= 6, \\
\Gamma(n) &= (n-1)!; \quad n = 1, 2, \ldots, \\
\Gamma(\alpha+1) &= \alpha\,\Gamma(\alpha); \quad \alpha > 0.
\end{aligned}$$

Para propiedades de la función gamma, véase su sección en el Apéndice.

La función de densidad (3.45) se escribe alternativamente como

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}; \quad x > 0, \quad \alpha, \lambda > 0, \tag{3.47}$$

donde $\lambda = 1/\beta > 0$ es un parámetro tasa. Esta función es no negativa e integra uno. De hecho, al considerar el cambio de variable $u = \lambda x$, se tiene

$$\begin{aligned}
\int_{-\infty}^\infty f(x)\mathrm{d}x = \int_0^\infty f(x)\mathrm{d}x &= \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}\mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)}\int_0^\infty (\lambda x)^{\alpha-1}e^{-\lambda x}(\lambda\mathrm{d}x) = \frac{1}{\Gamma(\alpha)}\int_0^\infty u^{\alpha-1}e^{-u}\mathrm{d}u \\
&= \frac{1}{\Gamma(\alpha)}\,\Gamma(\alpha) = 1.
\end{aligned}$$

El parámetro α modula la forma de la función de densidad. Destacan cuatro casos: (a) $0 < \alpha < 1$. En este caso, la población está muy concentrada a la derecha de su moda cero. De hecho, su función de densidad no está acotada superiormente:

$$f(0+) = \lim_{x\to 0+} f(x) = \lim_{x\to 0+}\left\{\frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}e^{-x/\beta}\right\} = \frac{1}{\Gamma(\alpha)\beta^\alpha}\lim_{x\to 0+} x^{\alpha-1} = \infty.$$

(b) $\alpha = 1$. Este es el caso de la distribución exponencial:

$$X \sim \text{gamma}(\alpha = 1, \beta) \sim \exp(\mu = \beta) \sim \exp(\lambda = 1/\beta).$$

Como el caso anterior, la función de densidad es estrictamente decreciente en $(0, \infty)$, aunque acotada superiormente. Por lo que su moda es cero. Además

$$f'(0+) < 0 < \lambda = f(0+).$$

(c) $1 < \alpha \leq 2$. En este caso la moda es positiva. Por la forma de la función de densidad, la población aún se aprecia relativamente cercana a cero. Aquí

$$f(0+) = 0 < f'(0+).$$

(d) $\alpha > 2$. Este caso es similar al anterior, en donde la población se aleja aún más del cero; lo que se refleja con una mayor moda. Además

$$f(0+) = f'(0+) = 0.$$

Por el Ejercicio 2.10, la moda es

$$\mathrm{moda}(X) = \text{máx}(\alpha - 1, 0)\beta = \frac{\text{máx}(\alpha - 1, 0)}{\lambda} \geq 0.$$

La mediana es el único punto $m > 0$ tal que

$$\int_0^m \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-x/\beta} \mathrm{d}x = \int_0^{m/\beta} \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-x} \mathrm{d}x = \frac{1}{2}.$$

Como $\beta > 0$ es un parámetro de escala, entonces

$$X \sim \mathrm{gamma}(\alpha, \beta) \quad \Leftrightarrow \quad cX \sim \mathrm{gamma}(\alpha, c\beta) \quad \Leftrightarrow \quad \frac{X}{\beta} \sim \mathrm{gamma}(\alpha, 1); \quad \text{con} \quad c > 0. \tag{3.48}$$

Por lo cual, la familia de distribuciones gamma es invariante ante un cambio de escala. Se verificará la última equivalencia. Si $X \sim \mathrm{gamma}(\alpha, \beta)$ y $Y = X/\beta$, entonces

$$F(y) = P(Y \leq y) = P(X/\beta \leq y) = P(X \leq \beta y) = F_X(\beta y); \quad \text{para} \quad y > 0.$$

Al derivar respecto de $y > 0$, se tiene

$$f(y) = \frac{\mathrm{d}}{\mathrm{d}y} F(y) = \frac{\mathrm{d}}{\mathrm{d}y}\{F_X(\beta y)\} = f_X(\beta y)\beta = \beta \frac{(\beta y)^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-\beta y/\beta} = \frac{y^{\alpha-1}}{\Gamma(\alpha)} e^{-y}.$$

La distribución gamma satisface

$$\mathrm{E}\, X = \alpha\beta, \tag{3.49}$$
$$\mathrm{V}\, X = \alpha\beta^2, \tag{3.50}$$
$$\mathrm{d.\,e.}(X) = \sqrt{\alpha}\beta, \tag{3.51}$$
$$\mathrm{sesgo}(X) = \frac{2}{\sqrt{\alpha}}, \tag{3.52}$$
$$\mathrm{curtosis}(X) = \frac{6}{\sqrt{\alpha}} + 3. \tag{3.53}$$

Se verificará lo anterior con la técnica de la función generadora de momentos

$$M(t) = \mathrm{E}\, e^{tX} = \frac{1}{(1-\beta t)^\alpha} = \frac{1}{(1-t/\lambda)^\alpha} = \left(\frac{\lambda}{\lambda - t}\right)^\alpha; \quad t < \lambda.$$

Esta fórmula se comprueba por lo siguiente

$$\begin{aligned} M(t) = \mathrm{E}\, e^{tX} &= \int_0^\infty e^{tx} f(x)\mathrm{d}x \\ &= \int_0^\infty e^{tx}\frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x}\mathrm{d}x = \lambda^\alpha \int_0^\infty \frac{x^{\alpha-1}}{\Gamma(\alpha)}e^{-(\lambda-t)x}\mathrm{d}x \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha}\int_0^\infty \frac{(\lambda-t)^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-(\lambda-t)x}\mathrm{d}x = \left(\frac{\lambda}{\lambda-t}\right)^\alpha; \quad t<\lambda. \end{aligned}$$

La última integral es uno, pues su integrando es tal cual la función de densidad gamma de parámetros forma α y tasa $\lambda - t > 0$. En cambio, para $t \geq \lambda$, la variable aleatoria e^{tX} tiene esperanza infinita y la función generadora de momentos no está definida. Así

$$\begin{aligned} M(t) &= (1-\beta t)^{-\alpha}, \\ M'(t) &= \alpha\beta(1-\beta t)^{-\alpha-1}, \\ M''(t) &= \alpha(\alpha+1)\beta^2(1-\beta t)^{-\alpha-2}, \\ M^{(3)}(t) &= \alpha(\alpha+1)(\alpha+2)\beta^3(1-\beta t)^{-\alpha-3}. \end{aligned}$$

Por lo que

$$\begin{aligned} \mathrm{E}\, X &= M'(0) = \alpha\beta, \\ \mathrm{E}\, X^2 &= M''(0) = \alpha(\alpha+1)\beta^2, \\ \mathrm{E}\, X^3 &= M^{(3)}(0) = \alpha(\alpha+1)(\alpha+2)\beta^3. \end{aligned}$$

Así

$$\begin{aligned} \mathrm{V}\, X = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2 &= \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 = \alpha\beta^2, \\ \mathrm{d.\,e.}(X) &= \sqrt{\alpha\beta^2} = \sqrt{\alpha}\beta. \end{aligned}$$

Para el cálculo del sesgo, considere $\beta = 1$, pues dicho coeficiente es invariante ante una transformación de escala. Por (2.77), se tiene

$$\begin{aligned} \mathrm{E}(X - \mathrm{E}\, X)^3 &= \mathrm{E}\, X^3 - 3\,\mathrm{E}\, X \cdot \mathrm{E}\, X^2 + 2(\mathrm{E}\, X)^3 \\ &= \alpha(\alpha+1)(\alpha+2) - 3\alpha\cdot\alpha(\alpha+1) + 2\alpha^3 \\ &= \alpha\left[(\alpha^2+3\alpha+2) - 3(\alpha^2+\alpha) + 2\alpha^2\right] \\ &= 2\alpha \end{aligned}$$

y

$$\mathrm{sesgo}(X) = \mathrm{E}\left(\frac{X-\mu}{\sigma}\right)^3 = \mathrm{E}\left(\frac{X-\alpha}{\sqrt{\alpha}}\right)^3 = \frac{\mathrm{E}(X-\alpha)^3}{\alpha^{3/2}} = \frac{2\alpha}{\alpha^{3/2}} = \frac{2}{\sqrt{\alpha}}.$$

Note que el coeficiente de sesgo tiende a anularse conforme α crece. Por último, del Ejercicio 2.12, se verifica que el coeficiente de curtosis es

$$\mathrm{curtosis}(X) = \frac{6}{\alpha} + 3.$$

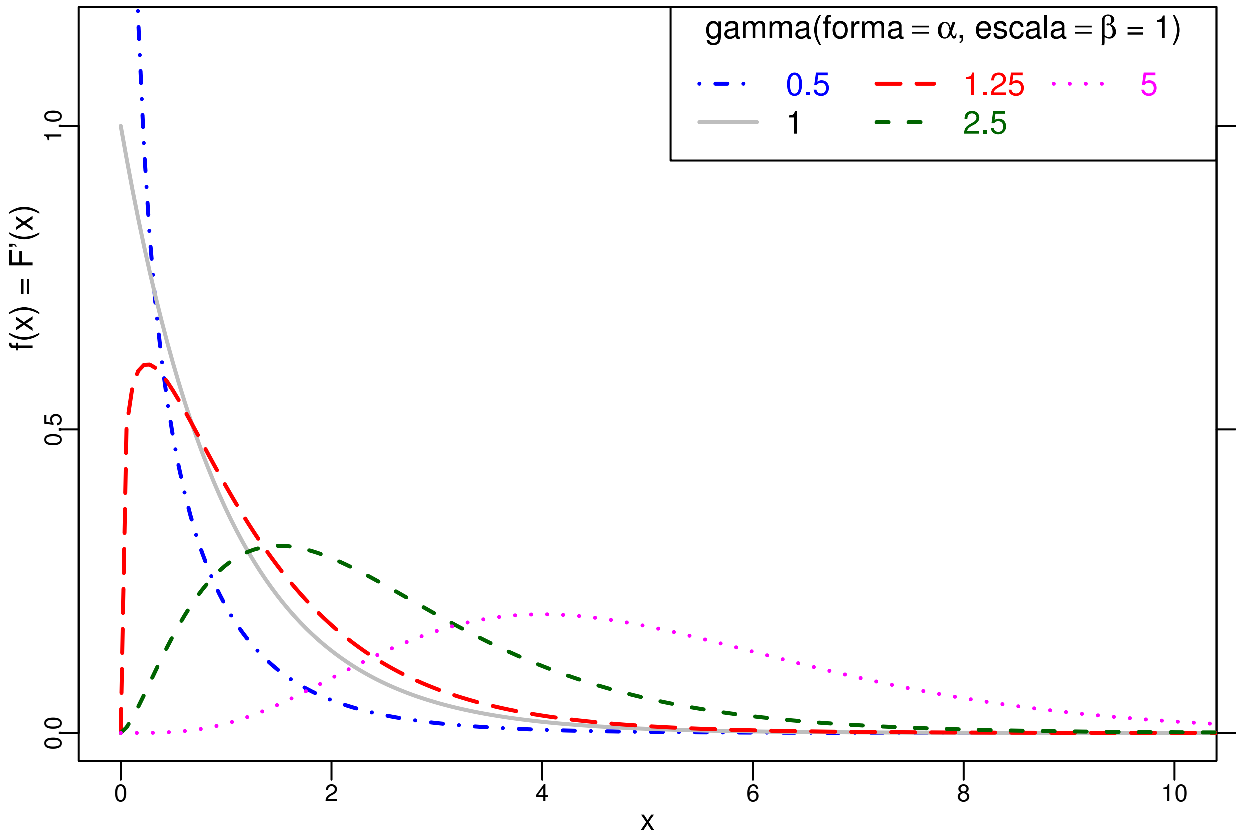


Figura 3.8. Gráficas de funciones de densidad gamma($\alpha, \beta = 1$), para diversos valores de α.

La curtosis tiende a 3, cuando α lo hace a ∞. Se concluye que la función de densidad gamma tiene una forma asimétrica con sesgo positivo. Conforme α crece, aumentan la media, mediana, moda y su dispersión (varianza y desviación estándar). En cambio, disminuyen su sesgo y curtosis. Son linealmente dependientes del parámetro escala β los indicadores poblacionales: media, mediana, moda y desviación estándar. Así mismo, el sesgo y la curtosis son invariantes ante β. En la Figura 3.8 se muestran ejemplos de gráficas de la función de densidad gamma, con $\alpha > 0$ y $\beta = 1$.

Distribución beta.* La distribución beta generaliza a la uniforme continua. Este es un modelo de probabilidad que aplica a poblaciones cuyos individuos toman cualquier valor entre dos números $-\infty < a < b < \infty$, con mayor frecuencia cerca de su moda; siempre que esta última sea única. Por simplicidad, el conjunto de valores posibles es $(a, b) = (0, 1)$. Una importante aplicación es en *estadística bayesiana*, donde la distribución a *priori* de cierto parámetro de interés es beta. En particular, la uniforme continua en $(0, 1)$ es una distribución a priori *no informativa*.

La función de densidad *beta* es

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}; \quad 0 < x < 1, \quad \alpha, \beta > 0, \tag{3.54}$$

donde la *función beta* es

$$\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)};\quad \alpha,\beta>0. \tag{3.55}$$

Recuerde que la función gamma se define en (3.46). La función de densidad beta integra uno, pues

$$\mathrm{B}(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x;\quad \alpha,\beta>0.$$

Por ejemplo, si $\alpha=2$ y $\beta=3$, la función de densidad beta es un polinomio de grado $\alpha+\beta-2=3$:

$$f(x)=\frac{\Gamma(2+3)}{\Gamma(2)\,\Gamma(3)}x^{2-1}(1-x)^{3-1}=\frac{4!}{1!2!}x(1-x)^2=12(x-2x^2+x^3);$$

para $0<x<1$. Así mismo

$$\begin{aligned}\int_{-\infty}^{\infty}f(x)\mathrm{d}x&=\int_0^1 f(x)\mathrm{d}x=\int_0^1 12(x-2x^2+x^3)\mathrm{d}x\\&=12\left[\frac{x^2}{2}-2\frac{x^3}{3}+\frac{x^4}{4}\right]_0^1=12\left[\frac{1}{2}-\frac{2}{3}+\frac{1}{4}\right]\\&=12\left[\frac{6-8+3}{12}\right]=1.\end{aligned}$$

Los parámetros $\alpha,\beta>0$ modulan la forma de la función de densidad beta. El primer parámetro perturba principalmente la forma de la cola izquierda mientras que el segundo lo hace en su cola derecha. La Figura 3.9[izquierda] muestra ejemplos de la gráfica de la función de densidad beta, para diversos valores de los parámetros $\alpha\neq\beta$. En este caso, la función de densidad es asimétrica. Si $1<\alpha<\beta$, la función de densidad está sesgada a la derecha, como se aprecia en los ejemplos $(\alpha,\beta)=(1.5,4.5),(2,3)$. En cambio, si $1<\beta<\alpha$, resulta un sesgo a la izquierda, como en los ejemplos $(\alpha,\beta)=(3,2),(4.5,2)$. De hecho, cambia el signo del sesgo al intercambiar el rol de los parámetros α y β. Por otra parte si $\alpha=\beta$, la función de densidad beta es simétrica al rededor de $x=1/2$. Este punto es la media, mediana y moda. Véanse las gráficas de funciones de densidad en la Figura 3.9[derecha], para los ejemplos $\alpha=\beta=0.5,1,2,4.5$. El caso $\alpha=\beta=1$ (curva horizontal) corresponde a la función de densidad uniforme en $(0,1)$. Note también que, cuando α y β crecen, la población es cada vez más homogénea al rededor de la media $\mu=1/2$. La función de densidad $f(x)$ es simétrica al rededor de dicho punto:

$$f\left(\frac{1}{2}-x\right)=f\left(\frac{1}{2}+x\right);\quad \text{para}\quad 0<x<\frac{1}{2}.$$

De hecho, si $\alpha=\beta$, se tiene

$$\left[\frac{1}{2}-x\right]^{\alpha}\left[1-\left(\frac{1}{2}-x\right)\right]^{\alpha}=\left[\frac{1}{2}-x\right]^{\alpha}\left[\frac{1}{2}+x\right]^{\alpha}=\left[\frac{1}{2}+x\right]^{\alpha}\left[1-\left(\frac{1}{2}+x\right)\right]^{\alpha}.$$

Por último, si $0<\alpha<1$ o $0<\beta<1$, entonces la función de densidad no está acotada superiormente en $x=0$ o $x=1$; respectivamente.

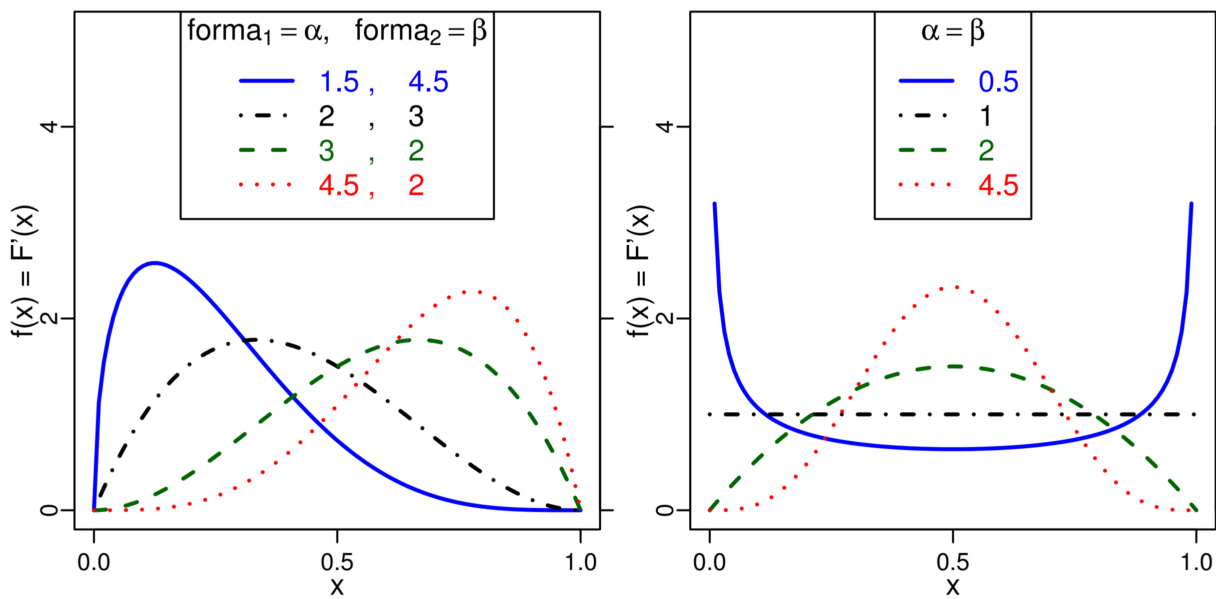


Figura 3.9. [izquierda] Gráficas de funciones de densidad beta: para $(\alpha, \beta) = (1.5, 4.5)$,$(2, 3)$,$(3, 2)$,$(4.5, 2)$. [derecha] Otros ejemplos, con $\alpha = \beta = 0.5, 1, 2, 4.5$.

Distribución de Cauchy. Una variable aleatoria X tiene distribución de *Cauchy*, si su función de densidad es

$$f(x) = \frac{1}{\pi\sigma\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}; \quad x \in \mathbf{R}, \quad \mu \in \mathbf{R}, \quad \sigma > 0. \tag{3.56}$$

Los parámetros μ y σ son de localización y escala; respectivamente. Note que esta distribución es simétrica al rededor de μ:

$$\begin{aligned} f(\mu - x) &= \frac{1}{\pi\sigma\left(1 + \left(\frac{(\mu-x)-\mu}{\sigma}\right)^2\right)} = \frac{1}{\pi\sigma\left(1 + \left(\frac{x}{\sigma}\right)^2\right)} \\ &= \frac{1}{\pi\sigma\left(1 + \left(\frac{(\mu+x)-\mu}{\sigma}\right)^2\right)} = f(\mu + x); \quad x \geq 0. \end{aligned}$$

Entonces, el parámetro de localización μ es tanto la moda como la mediana poblacional. Si embargo, esta población no tienen media. Para comprobar la última afirmación, considere el caso $\mu = 0$ y $\sigma = 1$:

$$\mathrm{E}\,|X| = \int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} \mathrm{d}x = 2\int_0^{\infty} \frac{x}{\pi(1+x^2)} \mathrm{d}x.$$

Aquí se usó la propiedad par de la función $x \to |x|/(1+x^2)$; $x \in \mathbf{R}$. Así

$$\mathrm{E}\,|X| = \int_0^{\infty} \frac{2x}{1+x^2} \mathrm{d}x = \int_1^{\infty} \frac{\mathrm{d}u}{u} = [\log u]_1^{\infty} = \log\infty - \log 1 = \infty.$$

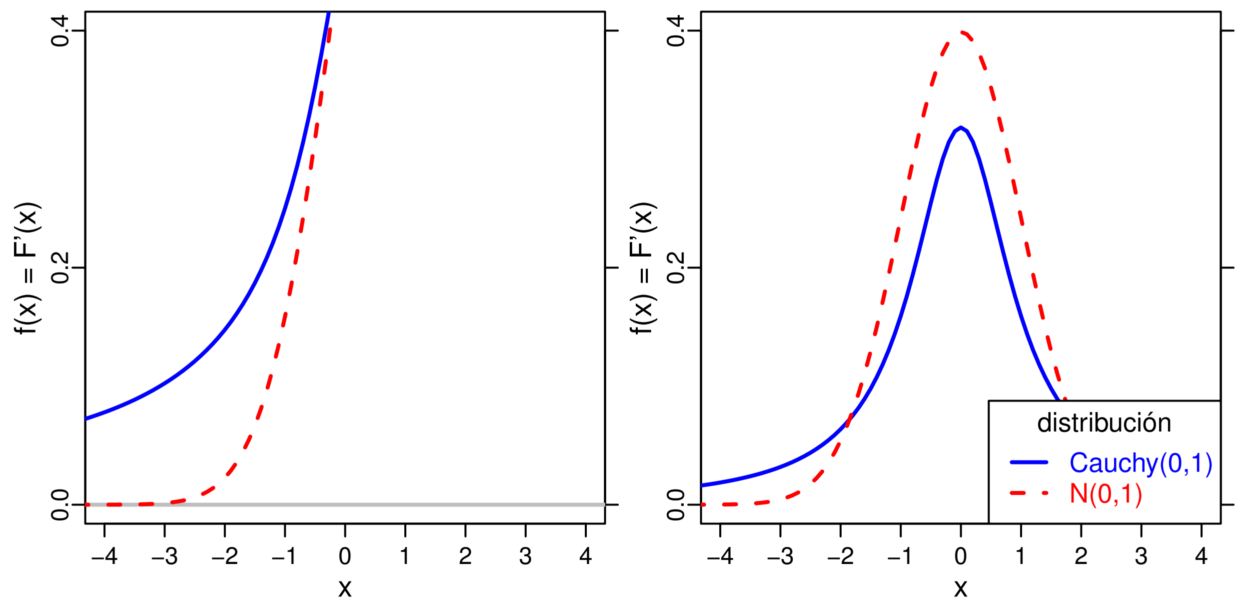


Figura 3.10. [I] Gráficas de función de densidad de Cauchy (gruesa) y de densidad normal estándar (delgada). [D] Correspondientes gráficas de funciones de distribución.

La Figura 3.10[izquierda] muestra la gráfica de la función de densidad de Cauchy($\mu = 0, \sigma = 1$), junto con la función de densidad normal estándar (3.24). Aquí se observa la forma simétrica al rededor de la mediana de la densidad de Cauchy. Además, respecto de una población normal, sus colas son extremadamente pesadas. Al no contar con media, la distribución de Cauchy no tiene momentos de orden $r \geq 1$. En consecuencia, tiene infinita varianza y curtosis. La función de distribución se obtiene como sigue:

$$
\begin{aligned}
F(x) = P(X \leq x) &= \int_{-\infty}^{x} f(t)\mathrm{d}t \\
&= \int_{-\infty}^{x} \frac{\mathrm{d}t}{\pi(1+t^2)} = \frac{1}{\pi}[\operatorname{atan} t]_{-\infty}^{x} \\
&= \frac{1}{\pi}\left[\operatorname{atan} x - \left(-\frac{\pi}{2}\right)\right] = \frac{1}{2} + \frac{1}{\pi}\operatorname{atan} x; \quad \text{para} \quad x \in \mathbf{R}.
\end{aligned}
$$

Recuerde que $(-\pi/2, \pi/2)$ es la imagen de la función $\operatorname{atan} x$. La Figura 3.10[derecha] muestra la gráfica de la función de distribución de Cauchy($\mu = 0, \sigma = 1$), junto con la de la distribución normal estándar (3.25). Observe que ambas funciones tienen una forma similar. Sin embargo, aquí se aprecia también que una población de distribución de Cauchy($\mu = 0, \sigma = 1$) es mucho más dispersa que la de una normal estándar.

Ejemplo 3.5.2. Sea X una variable aleatoria continua positiva, con función de densidad $f(x)$. Compruebe la *fórmula alternativa de Feller* de la esperanza

$$
\mathrm{E}\, X = \int_{0}^{\infty} P(X > x)\mathrm{d}x = \int_{0}^{\infty} (1 - F(x))\mathrm{d}x. \tag{3.57}
$$

Solución. Sea $f(x)$ la función de densidad de la variable aleatoria X. Como $X > 0$ con probabilidad uno, entonces

$$\begin{aligned}
\mathrm{E}\,X &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x = \int_{0}^{\infty} x f(x)\mathrm{d}x \\
&= \int_{0}^{\infty} \left(\int_{0}^{x} \mathrm{d}y\right) f(x)\mathrm{d}x = \int_{0}^{\infty}\int_{0}^{x} f(x)\mathrm{d}y\mathrm{d}x \\
&= \int_{0}^{\infty}\int_{y}^{\infty} f(x)\mathrm{d}x\mathrm{d}y = \int_{0}^{\infty} P(X > y)\mathrm{d}y.
\end{aligned}$$

La fórmula alternativa (3.57) es válida también para cualquier variable aleatoria no negativa; como discreta, continua, mezcla o de otro tipo; véase [2, Lemma V.6.1, p.p. 150].

Ejemplo 3.5.3. Sea X una variable aleatoria continua con mediana m, probar

$$\mathrm{E}\,|X - m| \leq \mathrm{E}\,|X - a|; \quad \text{para todo} \quad a \in \mathbf{R}.$$

Solución. Sin pérdida de generalidad considere que la mediana de la variable aleatoria X es cero. De hecho, si $Y = X - m$, entonces la mediana de Y es cero:

$$\begin{aligned}
P(Y \leq 0) &= P(X - m \leq 0) = P(X \leq m) = \frac{1}{2}, \\
P(Y \geq 0) &= P(X - m \geq 0) = P(X \geq m) = \frac{1}{2}.
\end{aligned}$$

Ahora se probará

$$\mathrm{E}\,|X| \leq \mathrm{E}\,|X - a|,$$

donde

$$\mathrm{E}\,|X| = \int_{-\infty}^{\infty} |x| f(x)\mathrm{d}x = -\int_{-\infty}^{0} x f(x)\mathrm{d}x + \int_{0}^{\infty} x f(x)\mathrm{d}x.$$

$$\begin{aligned}
& \mathrm{E}\,|X - a| \\
= \ & \int_{-\infty}^{\infty} |x - a| f(x)\mathrm{d}x \\
= \ & -\int_{-\infty}^{a} (x - a) f(x)\mathrm{d}x + \int_{a}^{\infty} (x - a) f(x)\mathrm{d}x \\
= \ & -\int_{-\infty}^{0} (x - a) f(x)\mathrm{d}x - \int_{0}^{a} (x - a) f(x)\mathrm{d}x + \int_{0}^{\infty} (x - a) f(x)\mathrm{d}x - \int_{0}^{a} (x - a) f(x)\mathrm{d}x \\
= \ & -\int_{-\infty}^{0} x f(x)\mathrm{d}x + \int_{0}^{\infty} x f(x)\mathrm{d}x + \int_{-\infty}^{0} a f(x)\mathrm{d}x - \int_{0}^{\infty} a f(x)\mathrm{d}x + 2\int_{0}^{a} (a - x) f(x)\mathrm{d}x \\
= \ & \mathrm{E}\,|X| + a\left(\int_{-\infty}^{0} f(x)\mathrm{d}x - \int_{0}^{\infty} f(x)\mathrm{d}x\right) + 2\int_{0}^{a} (a - x) f(x)\mathrm{d}x \\
= \ & \mathrm{E}\,|X| + a[P(X \leq 0) - P(X \geq 0)] + 2\int_{0}^{a} (a - x) f(x)\mathrm{d}x \\
= \ & \mathrm{E}\,|X| + a\left[\frac{1}{2} - \frac{1}{2}\right] + 2\int_{0}^{a} (a - x) f(x)\mathrm{d}x \\
\geq \ & \mathrm{E}\,|X|.
\end{aligned}$$

3.6. Transformación de variables aleatorias

Si X es una variable aleatoria definida en un espacio muestral Ω y $g : \mathbf{R} \to \mathbf{R}$ es una función real, entonces $Y = g(X)$ es también una variable aleatoria. Ejemplos de transformación de variables aleatorias son:

$$\begin{array}{lll} 4X+1, & X+c, & cX;\ \text{con } c \text{ una constante,} \\ e^X, & X^2, & \sqrt{X};\ \text{siempre que } P(X \geq 0) = 1, \\ X^3, & \operatorname{sen} X, & \dfrac{X}{X+1};\ \text{siempre que } P(X = -1) = 0. \end{array}$$

Por otro lado, hay transformaciones que involucran dos o más variables aleatorias. Si X y Y son variables aleatorias y a, b son constantes, algunos ejemplos de transformación son

$$aX + bY, \quad X \vee Y = \text{máx}(X, Y), \quad X \wedge Y = \text{mín}(X, Y), \quad \frac{X}{Y};\ \text{siempre que } P(Y = 0) = 0.$$

Cuando la función real $g(x)$ es continua y monótona creciente o decreciente, la distribución de la variable aleatoria $Y = g(X)$ se deduce fácilmente. Por ejemplo, considere

$$Y = e^X, \quad \text{con} \quad X \sim \mathrm{N}(0,1). \tag{3.58}$$

Entonces

$$\begin{aligned} F(y) &= P(Y \leq y) = P(e^X \leq y) \\ &= P(X \leq \log y) = \Phi(\log y); \quad y > 0, \end{aligned}$$

donde $\Phi(z)$ denota la función de distribución normal estándar (3.25):

$$\Phi(z) = \int_{-\infty}^{z} \phi(u)\mathrm{d}u = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\mathrm{d}u; \quad z \in \mathbf{R}.$$

Por (3.3) y (3.24), la función de densidad de Y es la derivada de su función de distribución:

$$\begin{aligned} f(y) = F'(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\Phi(\log y) \\ &= \Phi'(\log y)\frac{1}{y} = \frac{1}{y}\phi(\log y) \\ &= \frac{1}{\sqrt{2\pi}y} e^{-(\log y)^2/2}; \quad y > 0. \end{aligned}$$

Por lo tanto

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} e^{-(\log y)^2/2}; \quad y > 0. \tag{3.59}$$

Esta función de densidad es un caso particular de la distribución log normal, que se explica más adelante.

La Figura 3.11 muestra las gráfica de la transformación $y = g(x)$, junto con representaciones gráficas de la función de densidad normal estándar y su correspondiente función de densidad log normal. Al aplicar tal transformación, se aprecia el cambio de una distribución simétrica al rededor de cero a una con forma asimétrica de sesgo positivo.

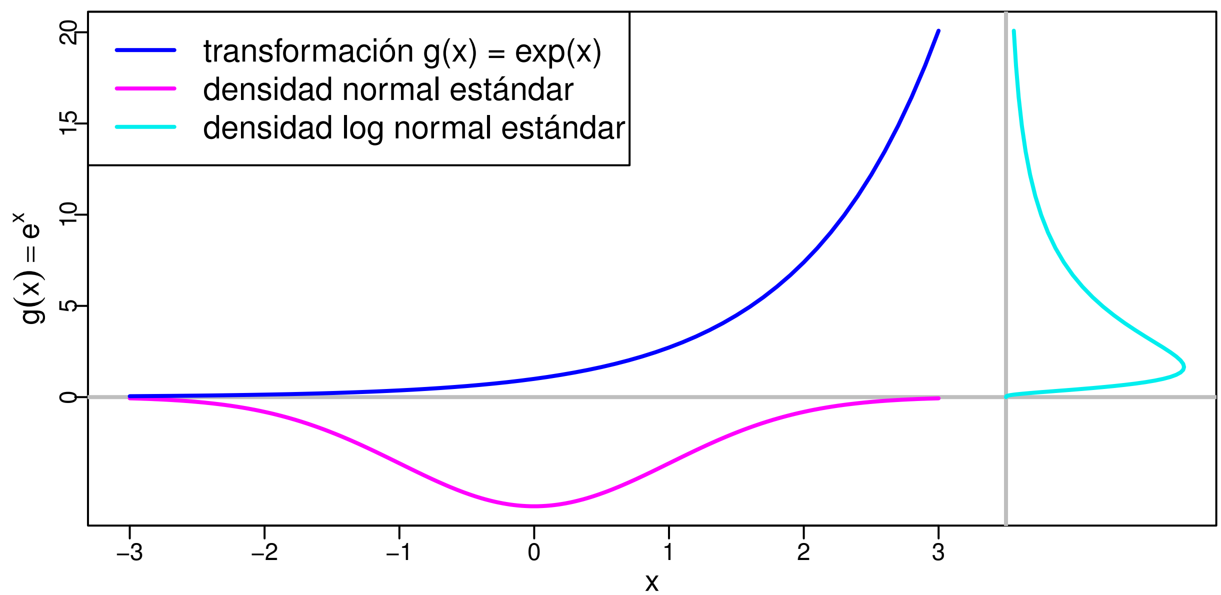


Figura 3.11. Gráficas de la transformación $g(x) = e^x$, la función de densidad normal estándar y su respectiva función de densidad log normal.

Teorema 3.6.1. Sean X una variable aleatoria continua, con conjunto de valores posibles $\mathcal{X}$ y función de densidad $f_X(x)$. Considere la variable aleatoria $Y = g(X)$, donde $g(x)$ es una función real derivable e invertible, definida en $\mathcal{X}$. Entonces, su función de densidad es

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|; \quad y \in \mathbf{R}. \tag{3.60}$$

Demostración. El conjunto de valores posibles de la variable aleatoria Y es

$$\mathcal{Y} = \{y = g(x) : x \in \mathcal{X}\}.$$

Como $g(x)$ es continua entonces, es una función estrictamente creciente o decreciente, al igual que su función inversa $h(y)$:

$$h(y) = x; \quad y \in \mathcal{Y} \quad \text{si y sólo si} \quad g(x) = y; \quad x \in \mathcal{X}.$$

Además $h(y)$ es derivable, con $h'(y) = 1/g'(h(y))$.
Caso $\underline{g(x) \text{ estrictamente creciente}}$. En este caso $h'(y) > 0$, para $y \in \mathcal{Y}$; casi en todas partes. Entonces

$$\{Y \leq y\} = \{g(X) \leq y\} = \{h(g(X)) \leq h(y)\} = \{X \leq h(y)\}$$

y

$$F_Y(y) = P(Y \leq y) = P(X \leq h(y)) = F_X(h(y)); \quad \text{para} \quad y \in \mathcal{Y}.$$

Al derivar respecto de y, se tiene

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(h(y)) \\ &= \frac{\mathrm{d}F_X}{\mathrm{d}y}(h(y)) \frac{\mathrm{d}}{\mathrm{d}y} h(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|; \quad y \in \mathcal{Y}. \end{aligned}$$

Caso $g(x)$ estrictamente decreciente. En este caso $h'(y) < 0$, para $y \in \mathcal{Y}$; casi en todas partes. Entonces

$$\{Y \le y\} = \{g(X) \le y\} = \{h(g(X)) \ge h(y)\} = \{X \ge h(y)\} = \{X < h(y)\}^c = \{X \le h(y)\}^c.$$

La última igualdad es *en probabilidad*, pues X es una variable aleatoria continua y $P(X = h(y)) = 0$. Así

$$F_Y(y) = P(Y \le y) = 1 - P(X \le h(y)) = 1 - F_X(h(y))$$

y

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}[1 - F_X(h(y))] \\ &= -\frac{\mathrm{d}}{\mathrm{d}y} F_X(h(y)) = \frac{\mathrm{d}F_X}{\mathrm{d}y}(h(y)) \left[-\frac{\mathrm{d}y}{\mathrm{d}h(y)} \right] \\ &= f_X(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|; \quad y \in \mathcal{Y}. \end{aligned}$$

$\square$

Este resultado se conoce como teorema de transformación. La fórmula invoca la regla de la cadena. Recuerde que una función de densidad es la derivada de su respectiva función de distribución. La función de densidad de Y se escribe también como

$$f_Y(y) = f_X(h(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} h(y) \right| = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|; \quad y \in \mathbf{R},$$

donde $h(y)$ denota la función inversa de $g(x)$:

$$x = h(y) = g^{-1}(y).$$

En el ejemplo de la distribución log normal (3.58)-(3.59), se tiene

$$\begin{aligned} y &= g(x) = e^x, \\ e^x &= y, \\ x &= \log y. \end{aligned}$$

La función inversa de $g(x)$ es

$$h(y) = g^{-1}(y) = \log y, \quad y > 0,$$

cuya derivada es

$$h'(y) = \frac{1}{y} > 0; \quad y > 0.$$

Por (3.60), del Teorema de transformación, se tiene (3.59):

$$\begin{aligned} f_Y(y) &= f_X(h(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}h(y)\right| = f_X(\log y)\left|\frac{\mathrm{d}}{\mathrm{d}y}\log y\right| \\ &= \frac{1}{\sqrt{2\pi}}e^{-(\log y)^2/2}\left|\frac{1}{y}\right| \\ &= \frac{1}{\sqrt{2\pi}\ y}e^{-(\log y)^2/2}; \quad y > 0. \end{aligned} \tag{3.61}$$

El conjunto de valores posibles de la variable aleatoria Y es

$$\mathcal{Y} = \{y = g(x) : x \in \mathcal{X}\} = \{y = e^x : x \in \mathbf{R}\} = \{y \in \mathbf{R} : y > 0\} = (0, \infty).$$

Ahora considere la transformación estrictamente decreciente $g(x) = e^{-x}$; para $x \in \mathbf{R}$. Entonces $Y = g(X) = e^{-X}$ y

$$\begin{aligned} y &= g(x) = e^{-x} > 0, \\ e^{-x} &= y, \\ -x &= \log y, \\ x &= -\log y, \\ \frac{\mathrm{d}x}{\mathrm{d}y} &= -\frac{1}{y} < 0. \end{aligned}$$

Por (3.60), del teorema de transformación, se tiene

$$\begin{aligned} f_Y(y) &= f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = f_X(-\log y)\left|-\frac{1}{y}\right| \\ &= \frac{1}{\sqrt{2\pi}}e^{-(-\log y)^2/2}\frac{1}{y} \\ &= \frac{1}{\sqrt{2\pi}\ y}e^{-(\log y)^2/2}; \quad y > 0. \end{aligned} \tag{3.62}$$

Observe que son iguales las funciones de densidad (3.61) y (3.62). Este resultado se explica también por (3.32):

$$\begin{aligned} -X &\overset{d}{=} X \sim \mathrm{N}(0,1), \\ e^{-X} &\overset{d}{=} e^X \sim \mathrm{logN}(0,1). \end{aligned}$$

En general, la exponencial de una variable aleatoria de distribución normal tiene distribución *log normal*:

$$X \sim \mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad e^X \sim \mathrm{logN}(\mu, \sigma^2),$$

Así mismo

$$Y \sim \log\mathrm{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad \log Y \sim \mathrm{N}(\mu, \sigma^2).$$

Al proceder como en (3.62), la función de densidad de la variable aleatoria $Y = e^X$ es

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left\{ -\frac{1}{2} \left(\frac{\log y - \mu}{\sigma} \right)^2 \right\}; \quad y > 0.$$

Cabe mencionar que los parámetros de la distribución log normal, $\mu \in \mathbf{R}$ y $\sigma^2 > 0$, no se refieren a la media o varianza de la variable aleatoria Y. De hecho e^μ es su mediana. Por (3.41) se obtiene tanto la media como la varianza de Y; vía la función generadora de momentos de la variable aleatoria X.

Por el Teorema de transformación, se verifica la propiedad de cerradura ante transformación de localización y escala de la familia de distribuciones normales (3.31). De hecho si

$$X \sim \mathrm{N}(\mu, \sigma^2) \quad \text{y} \quad Y = a + bX, \tag{3.63}$$

con $a, \mu \in \mathbf{R}$ y $b, \sigma^2 > 0$, entonces

$$\begin{aligned} y &= a + bx, \\ x &= \frac{y - a}{b}, \\ \frac{\mathrm{d}x}{\mathrm{d}y} &= \frac{1}{b} > 0 \end{aligned}$$

y

$$\begin{aligned} f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{y-a}{b} - \mu)^2/(2\sigma^2)} \left|\frac{1}{b}\right| \\ &= \frac{1}{\sqrt{2\pi}(b\sigma)} e^{-(\frac{y-a-b\mu}{b})^2/(2\sigma^2)} = \frac{1}{\sqrt{2\pi}(b\sigma)} e^{-(\frac{y-a-b\mu}{b\sigma})^2/2}. \end{aligned}$$

Por lo tanto, $Y \sim \mathrm{N}(a + b\mu, (b\sigma)^2)$.

Por otra parte, la familia de distribuciones uniforme continua es también cerrada bajo transformación de localización y escala:

$$X \sim \mathrm{U}(a, b) \quad \text{si y sólo si} \quad Y = c + dX \sim \mathrm{U}(c + ad, c + bd),$$

donde $-\infty < a < b < \infty$, $c \in \mathbf{R}$ y $d > 0$. Note que

$$\begin{aligned} a &< X < b, \\ c + ad &< c + dX < c + bd, \\ c + ad &< Y < c + bd \end{aligned}$$

y

$$\begin{aligned} y &= c + \mathrm{d}x, \\ x &= \frac{y - c}{d}, \\ \frac{\mathrm{d}x}{\mathrm{d}y} &= \frac{1}{d} > 0. \end{aligned}$$

Por (3.60), del Teorema de transformación, se tiene

$$f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = \frac{1}{b-a}\cdot\frac{1}{d} = \frac{1}{(b-a)d}; \quad \text{para} \quad c+ad < y < c+bd.$$

Por lo tanto, $Y \sim \mathrm{U}(c+ad, c+bd)$. Un resultado similar se obtiene si $b < 0$; véase Ejercicio 3.7.

Distribución Weibull. La distribución Weibull es una extensión de la distribución exponencial. Esta familia de distribuciones es utilizada en estadística industrial para determinar la confiabilidad de productos manufacturados.

Sea X una variable aleatoria de distribución exponencial de tasa $b > 0$. Defina la variable aleatoria $Y = X^{1/\alpha}$, con $\alpha > 0$. Esta es una transformación inducida por una función estrictamente creciente, con

$$\begin{aligned} g(x) &= x^{1/\alpha}, \quad x > 0, \\ x = h(y) &= y^{\alpha}, \\ h'(y) &= \alpha y^{\alpha-1} > 0, \quad y > 0. \end{aligned}$$

Por el teorema de transformación, la función de densidad de la variable aleatoria Y es

$$\begin{aligned} f(y) = f(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| &= be^{-b(y^{\alpha})}\cdot \alpha y^{\alpha-1} \\ &= \alpha b y^{\alpha-1}e^{-by^{\alpha}} = \alpha b^{1/\alpha}(b^{1/\alpha}y)^{\alpha-1}e^{-(b^{1/\alpha}y)^{\alpha}} \\ &= \alpha\lambda(\lambda y)^{\alpha-1}e^{-(\lambda y)^{\alpha}}; \quad y > 0, \end{aligned}$$

donde $\lambda = b^{1/\alpha} > 0$.

3.7. Simulación de variables aleatorias

Sea X una variable aleatoria con función de distribución $F(x)$. Su función *inversa generalizada*, *cuantil inferior* o *cuantil* es

$$F^{-}(\alpha) = \text{ínf}\{x \in \mathbf{R} : F(x) \geq \alpha\}; \quad \text{para} \quad \alpha \in (0,1). \tag{3.64}$$

En particular, cuando la función de distribución $F(x)$ es continua y estrictamente creciente en el conjunto de valores posibles (a, b), con $-\infty \leq a < b \leq \infty$, entonces esta es invertible, cuya inversa es el α-cuantil, α-cuantil inferior o cuantil α (3.39):

$$x_{\alpha} = F^{-1}(\alpha) = F^{-}(\alpha); \quad \alpha \in (0,1).$$

Por la propiedad de continuidad a la derecha de la función de distribución $F(x)$, se tiene

$$\{x \in \mathbf{R} : F(x) \geq \alpha\} = [F^{-}(\alpha), \infty). \tag{3.65}$$

Por lo que

$$F^{-}(\alpha) = \sup\{x \in \mathbf{R} : F(x) < \alpha\}; \quad \alpha \in (0,1). \tag{3.66}$$

Como $F(x)$ es creciente y continua por la derecha, entonces $F^{-}(\alpha)$ es función creciente y continua por la izquierda. Además

$$F^{-}(F(x)) \leq x; \quad \text{siempre que} \quad 0 < F(x) < 1. \tag{3.67}$$

De hecho

$$F^{-}(F(x)) = \text{ínf}\{y \in \mathbf{R} : F(y) \geq F(x)\} \leq x.$$

Así mismo, de la relación (3.65), se tiene

$$\alpha \leq F(F^{-}(\alpha)); \quad \alpha \in (0,1). \tag{3.68}$$

Teorema 3.7.1. Sea $F(x)$ una función de distribución con función inversa generalizada $F^{-}(\alpha)$. Defina la variable aleatoria X como

$$X = F^{-}(U), \quad \text{con} \quad U \sim \mathrm{U}(0,1).$$

Entonces, X tiene función de distribución $F(x)$.

Demostración. Es suficiente probar que

$$P(F^{-}(U) \leq x) = F(x). \tag{3.69}$$

Como $F(x)$ es función creciente, entonces

$$\begin{aligned} & P(F^{-}(U) \leq x) \\ \leq\ & P(F(F^{-}(U)) \leq F(x)) \\ \leq\ & P(U \leq F(x)) \\ =\ & F(x). \end{aligned}$$

La última desigualdad se debe a (3.68), aplicado como $U \leq F(F^{-}(U))$. Mientras que la última igualdad es porque $U \sim \mathrm{U}(0,1)$. De la misma manera, $F^{-}(\alpha)$ es función creciente y

$$\begin{aligned} F(x) &= P(U \leq F(x)) \\ &\leq P(F^{-}(U) \leq F^{-}(F(x))) \\ &\leq P(F^{-}(U) \leq x). \end{aligned}$$

La última desigualdad se debe a (3.67). Se concluye (3.69). □

Este resultado es importante para simular variables aleatorias con función de distribución conocida $F(x)$, puesto que se reduce a generar una variable aleatoria U con distribución uniforme en $(0,1)$ y evaluar en $F^{-}(U)$. Por otra parte, el siguiente teorema es un recíproco parcial en el caso de distribuciones continuas con función de distribución estrictamente creciente. En particular, este resultado incluye el caso absolutamente continuo.

Teorema 3.7.2. Sea X una variable aleatoria continua con conjunto de valores posibles un intervalo abierto (a, b); para ciertos $-\infty \leq a < b \leq \infty$, y función de distribución $F(x)$ estrictamente creciente en dicho intervalo. Entonces

$$F(X) \sim \mathrm{U}(0,1). \tag{3.70}$$

Recíprocamente, si $U \sim \mathrm{U}(0,1)$, entonces

$$F^{-1}(U) \stackrel{d}{=} X \sim F(x). \tag{3.71}$$

Demostración. Suponga que X es una variable aleatoria continua cuyo conjunto de valores posibles es un intervalo abierto:

$$\mathcal{X} = (a, b),$$

y que la función de distribución $F(x) = P(X \leq x)$; $x \in (a, b)$, es estrictamente creciente en dicho intervalo. Así, $F(x)$ es invertible y su función inversa satisface

$$F^{-1}(u) = x \quad \text{si y sólo si} \quad F(x) = u; \quad \text{para} \quad x \in \mathbf{R}, \quad u \in (0,1).$$

Entonces

$$\begin{aligned} P(F(X) \leq u) &= P(F^{-1}(F(X)) \leq F^{-1}(u)) = P(X \leq F^{-1}(u)) \\ &= F(F^{-1}(u)) = u; \quad \text{para} \quad 0 < u < 1. \end{aligned}$$

Por lo tanto, se obtiene (3.70).
El recíproco corresponde a un caso particular del Teorema 3.7.1. Sin embargo, con la hipótesis establecida, su demostración es más simple. Si $U \sim \mathrm{U}(0,1)$ y $a < x < b$, entonces

$$P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x).$$

Por lo que $F^{-1}(U)$ tiene función de distribución $F(x)$, es decir, se cumple (3.71). □

Este resultado se puede deducir del Teorema de transformación, siempre que su función de distribución $F(x)$ sea además derivable; véase el Ejercicio 3.8. Por ejemplo, si

$$X \sim \exp(\lambda) \quad \text{y} \quad U = F(X),$$

entonces $U \sim \mathrm{U}(0,1)$. Por (3.9), la función de distribución exponencial de tasa $\lambda > 0$ es

$$F(x) = 1 - e^{-\lambda x}; \quad x > 0.$$

Recíprocamente, si $U \sim \mathrm{U}(0,1)$, entonces

$$X = F^{-1}(U) \sim F(x).$$

Aquí la función inversa de $F(x)$ es

$$F^{-1}(u) = -\frac{1}{\lambda}\log(1-u); \quad 0 < u < 1.$$

distribución	valores posibles	función de distribución $F(x)$	inversa $F^{-1}(y)$
$\mathrm{U}(0,1)$	$(0,1)$	x	y
$\mathrm{U}(a,b)$	(a,b)	$\dfrac{x-a}{b-a}$	$a+y(b-a)$
$\exp(\mu)$	$(0,\infty)$	$1-e^{-x/\mu}$	$-\mu\log(1-y)$
$\mathrm{N}(0,1)$	$(-\infty,\infty)$	$\Phi(x)$	$\Phi^{-1}(y)$
$\mathrm{N}(\mu,\sigma^2)$	$(-\infty,\infty)$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	$\mu+\sigma\Phi^{-1}(y)$
$\log\mathrm{N}(\mu,\sigma^2)$	$(0,\infty)$	$\Phi\left(\frac{\log x-\mu}{\sigma}\right)$	$e^{\mu+\sigma\Phi^{-1}(y)}$

Tabla 3.3. Elementos base para la simulación de algunas distribuciones continuas.

La Tabla 3.3 muestra la función de distribución y su inversa de ciertas distribuciones de probabilidad continuas.

Por otra parte, si X es una variable aleatoria discreta, que toma valores en los enteros no negativos, entonces su función de distribución $F(x)$ es una función creciente escalonada, con salto en los enteros no negativos:

$$F(x) = F(\lfloor x\rfloor) = \sum_{i=0}^{\lfloor x\rfloor} f(i) = \begin{cases} 0 & x<0 \\ F(0) & 0\le x<1 \\ F(1) & 1\le x<2 \\ & \vdots \\ F(n) & n\le x<n+1 \\ & \vdots \end{cases}$$

para todo $n\geq 1$. Así, la función inversa generalizada queda: para $\alpha\in(0,1)$

$$F^{-1}(\alpha) = \begin{cases} 0 & 0<\alpha\le F(0) \\ 1 & F(0)<\alpha\le F(1) \\ 2 & F(1)<\alpha\le F(2) \\ \vdots & \\ n & F(n-1)<\alpha\le F(n) \\ \vdots & \end{cases}$$

De esta manera, $F^{-1}(\alpha)$ recupera a cada uno de los valores posibles de la variable aleatoria X. Note que $F^{-1}(\alpha)=n$, para α en un intervalo de longitud igual a la probabilidad del evento $\{X=n\}$:

$$f(n) = P(X=n) = F(n)-F(n-1);\quad n\geq 1.$$

Ahora considere una variable aleatoria U de distribución uniforme continua en el intervalo $(0,1)$ y defina $Y=F^{-1}(U)$. Del Teorema 3.7.1, se tiene que X y Y tienen la misma distribución. De hecho

$$\begin{aligned}P(Y=n) &= P(F^{-1}(U)=n) = P(F(n-1)<U\le F(n))\\ &= F(n)-F(n-1) = f(n)\\ &= P(X=n).\end{aligned}$$

Note además que la función inversa $F^{-}(\alpha)$ es continua por la izquierda y tiene límite por la derecha.

3.7.1. Desigualdades de Markov y Chevishev.

Las desigualdades de Markov y de Chebyshev relacionan la probabilidad de eventos asociados a una variable aleatoria con su valor esperado. Estas desigualdades son equivalentes, pues una lleva a la otra y viceversa. La desigualdad de *Markov* afirma que si Y es una variable aleatoria no negativa, es decir $P(Y\ge 0)=1$, entonces

$$P(Y>\varepsilon)\le \frac{\mathrm{E}\,Y}{\varepsilon};\quad \varepsilon>0. \tag{3.72}$$

Esta desigualdad, que aplica también cuando $\mathrm{E}\,Y=\infty$, se verificará para los casos discreto y continuo. Para el caso discreto, recuerde que

$$P(Y>\varepsilon)=\sum_{y>\varepsilon} f(y)\quad \text{y}\quad \mathrm{E}\,Y=\sum_{y} yf(y).$$

Luego

$$\begin{aligned}\mathrm{E}\,Y &= \sum_{y} yf(y) \ge \sum_{y>\varepsilon} yf(y)\\ &\ge \sum_{y>\varepsilon}\varepsilon f(y) = \varepsilon\sum_{y>\varepsilon} f(y)\\ &= \varepsilon P(Y>\varepsilon).\end{aligned}$$

De manera similar, en el caso continuo, recuerde que

$$P(Y>\varepsilon)=\int_{\varepsilon}^{\infty} f(y)\mathrm{d}y\quad \text{y}\quad \mathrm{E}\,Y=\int_{0}^{\infty} yf(y).$$

Por lo que

$$\begin{aligned}\mathrm{E}\,Y &= \int_{-\infty}^{\infty} yf(y)\mathrm{d}y = \int_{0}^{\infty} yf(y)\mathrm{d}y\\ &\ge \int_{\varepsilon}^{\infty} yf(y)\mathrm{d}y \ge \int_{\varepsilon}^{\infty}\varepsilon f(y)\mathrm{d}y\\ &= \varepsilon\int_{\varepsilon}^{\infty} f(y)\mathrm{d}y = \varepsilon P(Y>\varepsilon).\end{aligned}$$

Equivalente a la desigualdad de Markov, la desigualdad de *Chebyshev* es

$$P(|X-\mu|>\varepsilon) \le \frac{\mathrm{E}\,|X-\mu|^k}{\varepsilon^k}; \quad \varepsilon, k>0,$$

donde X es una variable aleatoria y $\mu = \mathrm{E}\,X \in \mathbf{R}$. De hecho, considere (3.72) con $Y = |X-\mu|^k$:

$$P(|X-\mu|>\varepsilon) = P(|X-\mu|^k > \varepsilon^k) \le \frac{\mathrm{E}\,|X-\mu|^k}{\varepsilon^k}; \quad \varepsilon>0.$$

En particular, para $k=1,2$, se tiene

$$P(|X-\mu|>\varepsilon) \le \frac{\mathrm{E}\,|X-\mu|}{\varepsilon}$$

y

$$P(|X-\mu|>\varepsilon) \le \frac{\mathrm{E}(X-\mu)^2}{\varepsilon^2} = \frac{\mathrm{V}\,X}{\varepsilon^2}; \quad \varepsilon>0.$$

3.7.2. Distribución del cociente

En esta sección se desarrollará la función de densidad del cociente de dos variables aleatorias continuas.

Considere dos variables aleatorias continuas X, Y, con función de densidad conjuntas $f(x,y)$; para $(x,y) \in \mathbf{R}$. Sea $Z = Y/X$. Esta es una variable aleatoria bien definida, pues $P(X=0)=0$. Para $z>0$, se tiene

$$\{Z \le z\} = \left\{\frac{Y}{X} \le z\right\} = \{X>0, Y \le zX\} \cup \{X<0, Y \ge zX\}.$$

Esta es una igualdad de conjuntos, casi en todas partes. Entonces

$$\begin{aligned} P(Z \le z) &= P(X>0, Y\le zX) + P(X<0, Y \ge zX) \\ &= \int_0^\infty \int_{-\infty}^{zx} f(x,y)\mathrm{d}y\mathrm{d}x + \int_{-\infty}^0 \int_{zx}^\infty f(x,y)\mathrm{d}y\mathrm{d}x. \end{aligned}$$

Al derivar respecto de z, la función de densidad de Z es

$$\begin{aligned} f(z) &= \mathrm{d}z P(Z \le z) \\ &= \int_0^\infty \mathrm{d}z \int_{-\infty}^{zx} f(x,y)\mathrm{d}y\mathrm{d}x + \int_{-\infty}^0 \mathrm{d}z \int_{zx}^\infty f(x,y)\mathrm{d}y\mathrm{d}x \\ &= \int_0^\infty [f_{X,Y}(x,zx)x]\mathrm{d}x + \int_{-\infty}^0 [-f_{X,Y}(x,zx)x]\mathrm{d}x \\ &= \int_{-\infty}^\infty |x| f_{X,Y}(x,zx)\mathrm{d}x. \end{aligned}$$

Se concluye

$$f(z) = \int_{-\infty}^\infty |x| f_{X,Y}(x,zx)\mathrm{d}x; \quad \text{para} \quad z \in \mathbf{R}. \tag{3.73}$$

Ejemplo 3.7.3. Sean X y Y dos variables aleatorias independientes de distribución gamma, $X \sim \text{gamma}(\alpha_1, \lambda)$, $Y \sim \text{gamma}(\alpha_2, \lambda)$, con $\alpha_1, \alpha_2, \lambda > 0$. Entonces, la función de densidad del cociente $Z = Y/X$ es

$$f(z) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)} \frac{z^{\alpha_2 - 1}}{(1+z)^{\alpha_1 + \alpha_2}}; \quad z > 0.$$

Por la fórmula (3.73), para $z > 0$, resulta

$$\begin{aligned}
f(z) &= \int_{-\infty}^{\infty} |x| f_{X,Y}(x, zx)\mathrm{d}x = \int_0^{\infty} x f(x) f_Y(zx)\mathrm{d}x \\
&= \int_0^{\infty} x \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1 - 1} e^{-\lambda x} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} (zx)^{\alpha_2 - 1} e^{-\lambda z x}\mathrm{d}x \\
&= \frac{\lambda^{\alpha_1 + \alpha_2} z^{\alpha_2 - 1}}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)} \int_0^{\infty} x^{\alpha_1 + \alpha_2 - 1} e^{-\lambda(1+z)x}\mathrm{d}x \\
&= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)} \frac{z^{\alpha_2 - 1}}{(1+z)^{\alpha_1 + \alpha_2}} \int_0^{\infty} \frac{[\lambda(1+z)]^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} x^{\alpha_1 + \alpha_2 - 1} e^{-\lambda(1+z)x}\mathrm{d}x \\
&= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)} \frac{z^{\alpha_2 - 1}}{(1+z)^{\alpha_1 + \alpha_2}}.
\end{aligned}$$

El último integrando es la función de densidad gamma de parámetros forma $\alpha_1 + \alpha_2$ y tasa $\lambda(1+z)$. Por lo que integra uno. Note que la distribución del cociente no tiene información del parámetro tasa λ. Si en particular $\lambda = 1/2$

Al comparar esta función de densidad con la de la distribución $\mathrm{F}(\alpha_1, \alpha_2)$, se tiene

$$\frac{\alpha_1}{\alpha_2} Z = \frac{Y/\alpha_2}{X/\alpha_1} \sim \mathrm{F}(\alpha_1, \alpha_2) \tag{3.74}$$

tiene distribución F, con α_1 y α_2 grados de libertad. Véase Ejercicio 3.22.

3.8. Ejercicios

Esta es la sección de ejercicios del capítulo de variables aleatorias continuas.

1. Sea X una variable aleatoria continua, con función de densidad $f(x) = F'(x)$; para $x \in \mathbf{R}$. Asuma que $f(x) > 0$, en cierto intervalo abierto (a, b): con $-\infty \leq a < b \leq \infty$. Compruebe que la función $F(x)$ es estrictamente creciente en dicho intervalo.

2. Sea X una variable aleatoria continua con función de densidad $f(x) = F'(x)$; para $x \in \mathbf{R}$. Demuestre la equivalencia de las fórmulas (3.4) y (3.2).

3. a) Compruebe que una variable aleatoria de distribución exponencial no tiene memoria:

$$P(X > x + y \mid X > x) = P(X > y); \quad x, y \geq 0. \tag{3.75}$$

   b) Recíprocamente, si X es una variable aleatoria continua, con conjunto de valores posibles $(0, \infty)$, que satisface la propiedad de pérdida de memoria (3.75), entonces X tiene distribución exponencial.

4. Considere el experimento de lanzar al azar un dardo en un disco de radio $r > 0$. El dardo cae con seguridad en cualquier punto del interior del disco. La probabilidad de caer en una región del interior del disco es proporcional a su área

$$P(A) = \frac{\text{área}(A)}{\pi r^2},$$

   donde A es cualquier conjunto de la σ-álgebra de Borel en el disco

$$\Omega = \{(u, v) \in \mathbf{R}^2 : u^2 + v^2 < r\}.$$

   Registre la distancia X del punto del dardo al centro. Luego, X es una variable aleatoria continua que toma cualquier valor en el intervalo $(0, r)$. La Figura 3.12[izquierda] muestra el círculo de radio $r = 0.35\,\text{m}$, así como la región B, acotada por los círculos concéntricos de radios $a = 0.2\,\text{m}$ y $b = 0.3\,\text{m}$. En su parte [derecha] se aprecia la gráfica de la función de distribución $F(x)$ de la mencionada variable aleatoria X.

   a) Demuestre que la función de distribución de X es

$$P(X \leq x) = \frac{x^2}{r^2}; \quad \text{para} \quad 0 \leq x \leq r$$

   b) Obtenga la gráfica de la función de distribución $F(x)$ y verifique su forma continua. Compare su resultado con la gráfica de la Figura 3.12[derecha].

   c) Para $0 \leq a \leq b \leq r$, compruebe

$$P(a < X \leq b) = \frac{b^2 - a^2}{r^2}$$

   d) Obtenga además la probabilidad de los siguientes eventos $X = 0$, $X = r$, $X = x$; para $x \in \mathbf{R}$ y $0 < X \leq r$

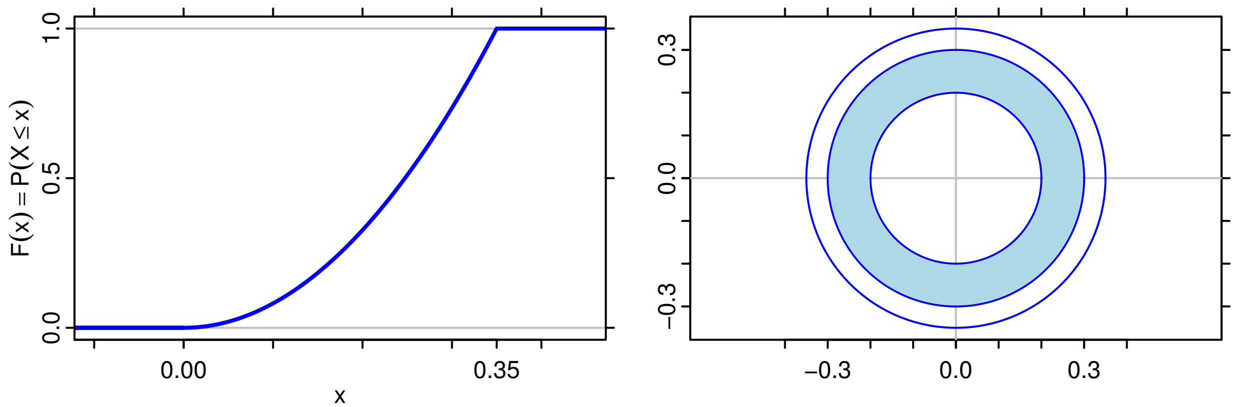


Figura 3.12. [izquierda] Círculo de radio $r = 0.35\,\text{m}$ y región concéntrica B, determinada por los radios $a = 0.2\,\text{m}$ y $b = 0.3\,\text{m}$. [derecha] Correspondiente gráfica de la función de distribución de la variable aleatoria X; Ejercicio 34.

e) Obtenga la función de densidad $f(x)$, así como su gráfica

f) Para $r = 1$, identifique la familia de distribuciones de probabilidad a la que pertenece la de la variable aleatoria X.

5. Compruebe que la distribución uniforme continua en el intervalo $(0, 1)$, es un caso particular de la distribución beta.

6. Se X una variable aleatoria de distribución uniforme continua: $X \sim \mathrm{U}(a = 0, b = 1)$. ¿qué distribución tiene la variable aleatoria $Y = c + dX$, cuando $d < 0$?

7. Considere las constantes $-\infty < a < b < \infty$, $c \in \mathbf{R}$ y $d < 0$. Con el teorema de transformación compruebe que, si la variable aleatoria X tiene distribución uniforme continua en el intervalo (a, b), entonces $Y = a + dX$, también tiene distribución uniforme continua. Sugerencia: obtenga primero los extremos de la variable aleatoria Y.

8. Con el Teorema de transformación, compruebe el Teorema 3.7.2, para el caso de una variable aleatoria continua X, con función de distribución $F(x)$:

$$U = F(X) \sim \mathrm{U}(0, 1), \quad \text{si y sólo si} \quad X = F^{-1}(U) \sim F(x).$$

Sugerencia: para el cálculo de la derivada de $F^{-1}(u)$, considere la fórmula

$$\frac{\mathrm{d}}{\mathrm{d}u} F^{-1}(u) = \frac{1}{F'(F^{-1}(u))}.$$

9. Considere una variable aleatoria X, con función de distribución

$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ 2x - x^2 & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

a) Escriba el conjunto de valores posibles de la variable aleatoria X. b) Obtenga la función de densidad $f(x)$. c) Obtenga la esperanza de X d) Obtenga la varianza de X. e) ¿Cuál es la probabilidad del evento $\{X \geq 0\}$.

10. Sea X una variable aleatoria de distribución gamma (3.45), de parámetros $\alpha, \beta > 0$. Para $\alpha > 1$, compruebe que la moda es

$$\mathrm{moda}(X) = (\alpha - 1)\beta > 0.$$

11. Sea X una variable aleatoria de distribución gamma (3.45), de parámetros $\alpha, \beta > 0$. Con el teorema de transformación, compruebe la primera equivalencia en (3.48). Por lo tanto, la familia de distribuciones gamma es invariante ante un cambio de escala.

12. Sea X una variable aleatoria de distribución gamma de parámetros $\alpha, \beta > 0$. Con la función generadora de momentos, compruebe que el coeficiente de curtosis es $6/\alpha + 3$. Sugerencia: considere el caso $\beta = 1$.

13. La longitud en centímetros del pez robalo o róbalo maduro en el mes de agosto es de una media poblacional de 73.4 y 86.4 para machos y hembras; respectivamente. Las correspondientes desviaciones estándar son 5.4 y 10.6 cm. Asuma que ambas poblaciones tienen distribución normal.

    a) En una sola figura, obtenga la gráfica de ambas funciones de densidad.

    b) Compare la media de ambas poblaciones, así como la desviación estándar.

    c) Obtenga un intervalo de predicción del 95 % para la longitud de los machos y hembras:

    $$\begin{aligned} X &\in \mu_1 \pm 1.96\sigma_1, \\ Y &\in \mu_2 \pm 1.96\sigma_2. \end{aligned}$$

    d) La relación macho-hembra es 1.6 : 1. La longitud de un pez elegido al azar es

    $$W = RX + (1 - R)Y,$$

    donde R es una variable aleatoria de distribución Bernoulli de parámetro

    $$p = \frac{1.6}{1.6 + 1} = \frac{1.6}{2.6} = 0.6154.$$

    Esta es una mezcla de dos poblaciones independientes de distribución normal, la cual tiene también distribución normal. Encuentre $\mathrm{E}\,W$, $\mathrm{E}\,W^2$, $\mathrm{V}\,W$ y $\mathrm{d.\,e.}(W)$.

    e) Obtenga una gráfica de la función de densidad de W. Compare su forma con la de las dos densidades del inciso a).

14. Sea X una variable aleatoria continua o discreta, con función de densidad $f(x)$. Demuestre que X es una variable aleatoria simétrica respecto de cierto $\mu \in \mathbf{R}$ si y sólo si $f(\mu - x) = f(\mu + x)$; para todo $x \in \mathbf{R}$.

15. Sea X una variable aleatoria continua, con función de distribución $F(x) = P(X \leq x)$; para $x \in \mathbf{R}$. Demuestre que X es una variable aleatoria simétrica respecto de cierto $\mu \in \mathbf{R}$ si y sólo si
$$F(\mu + x) = 1 - F(\mu - x); \quad \text{para todo} \quad x \in \mathbf{R}.$$

16. Obtenga un contra ejemplo de la identidad del inciso anterior, para una variable aleatoria simétrica X.

17. Sea $f(x)$ la función de densidad normal de media $\mu \in \mathbf{R}$ y varianza $\sigma^2 > 0$ (3.23). Compruebe que esta función es simétrica respecto de μ.

18. Sea X una variable aleatoria continua o discreta. Para los siguientes ejemplos solicitados, se puede apoyar en una gráfica o bosquejo.

    - Obtenga un ejemplo de función de densidad $f(x)$ que sea simétrica respecto de cero, tal que este punto no sea moda.
    - Obtenga un ejemplo de función de densidad $f(x)$ que sea simétrica respecto de cero, tal que este punto no sea moda ni mediana única, según la definición (2.36).
    - Obtenga un ejemplo de función de densidad $f(x)$ que sea simétrica respecto de cero, tal que este punto no sea moda ni mediana única, según (2.36), así como sin media definida.

19. Sea X una variable aleatoria continua con función de densidad $f(x) = F'(x)$; para $x \in \mathbf{R}$. Para los siguientes ejemplos solicitados, se puede apoyar en una gráfica o bosquejo.

    - Obtenga un ejemplo de densidad $f(x)$, tal que la mediana no sea única, según (2.36).
    - Si el conjunto de valores posibles de X es un intervalo abierto (a, b), con $-\infty \leq a < b \leq \infty$, tal que $f(x) > 0$ en dicho intervalo, compruebe que la mediana es única.

20. Sean X una variable aleatoria de distribución normal de media $\mu \in \mathbf{R}$ y varianza $\sigma^2 > 0$, $a \in \mathbf{R}$ y $b \neq 0$. Compruebe que la variable aleatoria $Y = a + bX$ tiene distribución normal de media $a + b\mu$ y varianza $b^2\sigma^2$. Sugerencia: para los casos $b > 0$ y $b < 0$, desarrolle la función de distribución $F_Y(y) = P(Y \leq y)$; para $y \in \mathbf{R}$.

21. Sea X una variable aleatoria de distribución normal de media $\mu \in \mathbf{R}$ y varianza $\sigma^2 > 0$. Por (3.40), su función generadora de momentos es $M(t) = Ee^{tX} = e^{\mu t + \sigma^2 t^2/2}$; $t \in \mathbf{R}$. Sus primera y segunda derivada son (3.42) y (3.43); respectivamente.

    - Compruebe
$$M^{(3)}(t) = [3\sigma^2(\mu + \sigma^2 t) + (\mu + \sigma^2 t)^3]M(t),$$
$$M^{(4)}(t) = [3\sigma^4 + 6\sigma^2(\mu + \sigma^2 t)^2 + (\mu + \sigma^2 t)^4]M(t).$$

    - Obtenga $\mathrm{E}\, X^3$ y $\mathrm{E}\, X^4$, así como confirme que sesgo$(X) = 0$ y curtosis$(X) = 3$. En consecuencia, el sesgo y la curtosis son constantes.

22. Considere el Ejemplo 3.73, de la función de densidad del cociente de dos variables aleatorias independientes de distribución gamma: $X \sim \text{gamma}(\alpha_1, \lambda)$, $Y \sim \text{gamma}(\alpha_2, \lambda)$ y $Z = Y/X$. Demuestre que $W = (\alpha_2/\alpha_1)Z$ tiene distribución F-Fisher, con α_1 y α_2 grados de libertad.

23. Para una población de distribución normal estándar, encontrar (a) Los primeros cuatro momentos centrales, y (b) Los coeficientes de sesgo y curtosis.

24. Sea $f(x)$ la función de densidad normal de media $\mu \in \mathbf{R}$ y varianza $\sigma^2 > 0$. Compruebe

    a) $f(x)$ es simétrica respecto de μ.

    b) $f(x)$ tiene un máximo global en $x = \mu$ y $f(\pm\infty) = 0$.

    c) Los puntos $x = \mu \pm \sigma$ son donde $f(x)$ cambia de convexa a cóncava o viceversa.

    d) Represente con una gráfica las propiedades anteriores.

25. Sea $\phi(z)$ la función de densidad normal estándar (3.24). Compruebe

$$\int_{-\infty}^{\infty} \phi(z) \mathrm{d}z = 1.$$

    Sugerencia: realice los siguientes pasos:

    a) Como $\phi(z)$ es una función par o simétrica respecto de $z = 0$, entonces

$$\int_{-\infty}^{\infty} \phi(z) \mathrm{d}z = \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} e^{-z^2/2} \mathrm{d}z.$$

    b) Con un cambio de variable, exprese la última integral en términos de $\int_0^\infty e^{-u^2} \mathrm{d}u$.

    c) Por último, mediante un cambio de coordenadas rectangulares a polares, resuelva la integral doble $\int_0^\infty \int_0^\infty e^{-(u^2+v^2)} \mathrm{d}v \mathrm{d}u$.

26. Sea $f(x)$ la función de densidad de Cauchy(μ, σ) (3.56). Confirme que esta función satisface sus dos propiedades características.

27. Sea X una variable aleatoria continua positiva, con función de densidad $f(x)$. Compruebe una generalización de la fórmula alternativa de la esperanza de Feller (3.57):

$$\mathrm{E}\, X^r = r \int_0^\infty x^{r-1} P(X > x) \mathrm{d}x = r \int_0^\infty x^{r-1} (1 - F(x)) \mathrm{d}x; \quad \text{para} \quad r \geq 1.$$

Capítulo 4

Vectores aleatorios continuos

Este capítulo es una introducción de los vectores aleatorios continuos. Se desarrollará esencialmente el caso bivariado; de dimensión dos.

En la Sección 2.4 se definió a un *vector aleatorio* de dimensión dos o bidimensional (X, Y), en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, como par de variables aleatorias X y Y. Como en el caso univariado, antes de que se lleve a cabo el experimento, su función de *distribución* o de *distribución conjunta* tiene toda la información el vector aleatorio (X, Y):

$$F(x, y) = P(X \leq x, Y \leq y); \quad (x, y) \in \mathbf{R}^2. \tag{4.1}$$

Por ejemplo, si

$$-\infty \leq a \leq b \leq \infty \quad \text{y} \quad -\infty \leq c \leq d \leq \infty,$$

entonces

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(b, c) - F(a, d) + F(a, c). \tag{4.2}$$

Esta es una fórmula análoga a la del caso unidimensional (2.5). Para verificarla, defina los rectángulos en el plano

$$\begin{aligned} B &= (a, b] \times (c, d], & C_1 &= (-\infty, a] \times (-\infty, d], \\ D &= (-\infty, b] \times (-\infty, d], & C_2 &= (-\infty, b] \times (-\infty, c]. \end{aligned}$$

Así

$$D = B \cup (C_1 \cup C_2), \quad \text{con} \quad B \cap (C_1 \cup C_2) = \varnothing.$$

La Figura 4.1 muestra la gráfica de los rectángulos B, C_1, C_2, $C_1 \cap C_2$, así como la unión D de todos ellos. Por lo que

$$\begin{aligned} F(b, d) &= P(X \leq b, Y \leq d) \\ &= P((X, Y) \in D) \\ &= P((X, Y) \in B \cup (C_1 \cup C_2)) \\ &= P((X, Y) \in B) + P((X, Y) \in C_1 \cup C_2), \\ &= P((X, Y) \in B) + P((X, Y) \in C_1) + P((X, Y) \in C_2) - P((X, Y) \in C_1 \cap C_2), \\ &= P((X, Y) \in B) + F(a, d) + F(b, c) - F(a, c). \end{aligned}$$

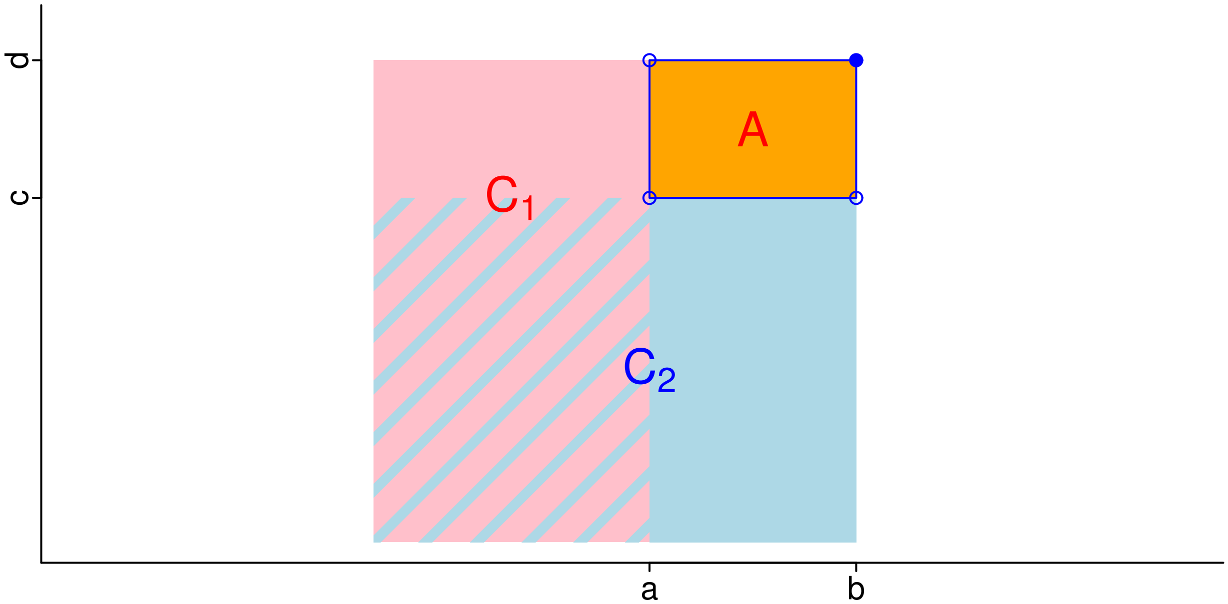


Figura 4.1. Gráfica de la región de integración de los cinco rectángulos involucrados en la fórmula (4.2).

Así

$$P(a < X \leq b, c < Y \leq d) = P((X,Y) \in B) = F(b,d) - F(a,d) - F(b,c) + F(a,c).$$

Se dice que el vector aleatorio (X, Y) es *continuo* si sus dos elementos son variables aleatorias continuas. Por el Ejercicio 4.2, esta condición equivale que sea continua su función de distribución conjunta $F(x, y)$. Además, (X, Y) es un vector aleatorio *absolutamente continuo* si su función de distribución es *absolutamente continua*:

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\mathrm{d}u\mathrm{d}v; \tag{4.3}$$

para $x, y \in \mathbf{R}$, donde $f(x, y)$ es la función de *densidad* o de *densidad conjunta* de (X, Y). Esta función es única casi en todas partes y satisface

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y) = \frac{\partial^2}{\partial y \partial x}F(x,y); \quad x, y \in \mathbf{R}, \quad \text{casi en todas partes.}$$

El *conjunto de vectores posibles* de (X, Y) es la región del plano donde la función de densidad conjunta esté definida y sea positiva:

$$\{(x,y) \in \mathbf{R}^2 : f(x,y) > 0\} \subset \mathbf{R}^2.$$

Por simplicidad, se omitirá la frase "casi en todas partes". Las propiedades características de una función de densidad continua bivariada son:

1.

$$f(x,y) \geq 0; \quad \text{para} \quad (x,y) \in \mathbf{R}^2. \tag{4.4}$$

2.
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = 1. \tag{4.5}$$

El intercambio de los límites de integración se debe al teorema de Tonelli Fubini. Recíprocamente, si existe una función real $f(x, y)$, que satisfaga las dos propiedades anteriores, entonces dicha función corresponde a la función de densidad conjunta de cierto vector aleatorio absolutamente continuo (X, Y). La función de densidad conjunta es la única función real no negativa tal que

$$P((X,Y)\in B) = \iint_B f(x,y)\mathrm{d}y\mathrm{d}x; \quad B \in \mathcal{B}(\mathbf{R}^2).$$

La probabilidad es el volumen de $f(x, y)$ bajo la superficie B. En particular, con el apoyo visual de la Figura 4.1, la probabilidad en (4.2) es

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x,y)\mathrm{d}y\mathrm{d}x = \int_c^d \int_a^b f(x,y)\mathrm{d}x\mathrm{d}y.$$

De nuevo, el intercambio de los límites de integración se debe al teorema de Tonelli Fubini. Así mismo, para (4.3), la región de integración aplicada es

$$B = (-\infty, x] \times (-\infty, y] = \{(u,v) \in \mathbf{R}^2 : u \leq x, v \leq y\}; \quad x, y \in \mathbf{R}. \tag{4.6}$$

Este rectángulo es la intersección de los rectángulos no acotados

$$\begin{aligned} (-\infty, x] \times (-\infty, \infty) &= \{(u,y) \in \mathbf{R}^2 : u \leq x\}, \\ (-\infty, \infty) \times (-\infty, y] &= \{(x,v) \in \mathbf{R}^2 : v \leq y\}. \end{aligned} \tag{4.7}$$

La Figura 4.2 muestra la gráfica de dichos rectángulos.

4.1. Distribución marginal y condicional

Sean X y Y dos variables aleatorias continuas con función de densidad conjunta $f(x, y)$. La función de *distribución marginal* de la variable aleatoria X es su correspondiente función de distribución en el contexto unidimensional:

$$F(x) = P(X \leq x) = P(X \leq x, Y < \infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,y)\mathrm{d}y\mathrm{d}u; \quad x \in \mathbf{R}. \tag{4.8}$$

Aquí, la región de integración es el rectángulo en el plano (4.7). Por otro lado, recuerde que se satisface

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u; \quad x \in \mathbf{R}, \tag{4.9}$$

donde $f(x)$ denota su función de *densidad marginal*:

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x); \quad x \in \mathbf{R}.$$

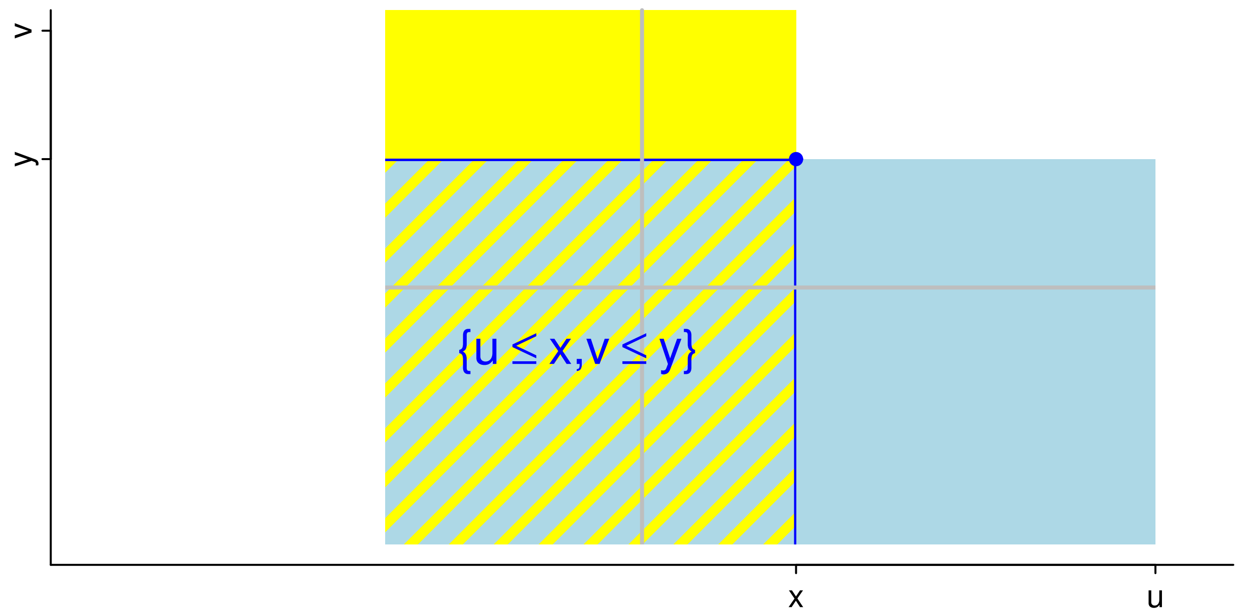


Figura 4.2. Gráfica de la región de integración de (4.6); el rectángulo $\{u \leq x, v \leq y\}$.

Al comparar los integrandos de (4.8) y (4.9), y por unicidad de la función de densidad, se tiene

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y; \quad u \in \mathbf{R}.$$

La función de densidad marginal de Y se obtiene de manera similar. Se concluye

$$f(x) = f_X(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y; \quad x \in \mathbf{R}, \tag{4.10}$$

$$f(y) = f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}x; \quad y \in \mathbf{R}. \tag{4.11}$$

Por otro lado, recuerde que las variables aleatorias X y Y son independientes, si se cumple (2.46):

$$F(x, y) = F(x)F(y); \quad x, y \in \mathbf{R}.$$

Análogo al caso discreto (2.47), la independencia entre las variables aleatorias continuas X y Y equivale a que la función de densidad conjunta sea el producto de sus densidades marginales:

$$f(x, y) = f(x)f(y); \quad \text{para todo} \quad x, y \in \mathbf{R}. \tag{4.12}$$

Véase Ejercicio 4.3.

Sea (X, Y) un vector aleatorio continuo con densidad conjunta $f(x, y)$ y considere una función real $g(x, y)$, tal que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g(x, y)| f(x, y)\mathrm{d}y\mathrm{d}x < \infty.$$

Entonces

$$\mathrm{E}\, g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\mathrm{d}y\mathrm{d}x. \tag{4.13}$$

Esta fórmula es consistente con la obtenida para el caso de variables aleatorias continuas univariadas (3.17). Por ejemplo, si $g(x,y) = x$; para $x, y \in \mathbf{R}$, se tiene

$$\mathrm{E}\, X = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} x\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x. \tag{4.14}$$

Así mismo, con $g(x,y) = y$, se tiene

$$\mathrm{E}\, Y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} y\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} yf(y)\mathrm{d}y. \tag{4.15}$$

La covarianza de X y Y (2.51), se obtiene con $g(x,y) = (x-\mu_1)(y-\mu_2)$ en (4.13), con $\mu_1 = \mathrm{E}\, X$ y $\mu_2 = \mathrm{E}\, Y$. Además, por (4.5), (4.14) y (4.15), se ratifica (2.53):

$$\begin{aligned}
\mathrm{C}(X,Y) = \mathrm{E}[(X-\mu_1)(Y-\mu_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)(y-\mu_2)f(x,y)\mathrm{d}y\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(xy-\mu_1 y-\mu_2 x+\mu_1\mu_2)f(x,y)\mathrm{d}y\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\mathrm{d}y\mathrm{d}x - \mu_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\mathrm{d}y\mathrm{d}x \\
&\quad - \mu_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\mathrm{d}y\mathrm{d}x + \mu_1\mu_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x \\
&= \mathrm{E}\, XY - \mu_1\mu_2 - \mu_1\mu_2 + \mu_1\mu_2 = \mathrm{E}\, XY - \mu_1\mu_2.
\end{aligned}$$

Por (2.58), el coeficiente de correlación no depende de la escala o unidad de medida de las variables aleatorias X y Y. De hecho, similar a los coeficientes de sesgo y curtosis, este coeficiente se define en términos de las versiones estandarizadas de X y Y.

Cabe recordar también que cero correlación no implica independencia; véase el Ejemplo 2.4.5. Otro contraejemplo de vector aleatorio continuo se obtiene como sigue. Considere dos variables aleatorias X y Y tal que

$$Y = X^2, \quad \text{con} \quad \mathrm{E}\, X = \mathrm{E}\, X^3 = 0.$$

Esta condición se obtiene en particular cuando X es una variable aleatoria simétrica al rededor de cero, como la distribución normal estándar. Entonces

$$\begin{aligned}
\mathrm{E}\, XY &= \mathrm{E}\, X(X^2) = \mathrm{E}\, X^3 = 0, \\
\mathrm{E}\, X \cdot \mathrm{E}\, Y &= 0 \cdot \mathrm{E}\, Y = 0, \\
\mathrm{C}(X,Y) &= \mathrm{E}\, XY - \mathrm{E}\, X \cdot \mathrm{E}\, Y = 0.
\end{aligned}$$

La variable aleatoria Y no aporta información del signo de X, aunque si de su valor absoluto. Si bien X y Y no son independientes, su información compartida se describe con el concepto de distribución condicional; que se define a continuación para el caso continuo.

Sean X y Y dos variables aleatorias continuas, con funciones de densidad conjunta $f(x,y)$ y marginales $f(x)$ y $f(y)$. La función de *densidad condicional* de la variable aleatoria Y, dado el evento $\{X = x\}$, es

$$f(y \mid x) = \frac{f(x,y)}{f(x)}; \quad \text{para} \quad x, y \in \mathbf{R}, \quad \text{con} \quad f(x) > 0. \tag{4.16}$$

En cambio, $f(y \mid x) = 0$, si $f(x) = 0$. Esta es una función de dos variables: (x, y). Sin embargo, cada variable tiene un rol diferente. La variable de interés es $y \in \mathbf{R}$, como valor posible de Y. Mientras que $x \in \mathbf{R}$ se considera fijo. La función $f(y \mid x)$ es una función de densidad respecto de y, pues satisface (3.7)-(3.8). De hecho, $f(y \mid x)$ es no negativa e

$$\int_{-\infty}^{\infty} f(y \mid x) \mathrm{d}y = \int_{-\infty}^{\infty} \frac{f(x,y)}{f(x)} \mathrm{d}y = \frac{1}{f(x)} \int_{-\infty}^{\infty} f(x,y) \mathrm{d}y = \frac{f(x)}{f(x)} = 1.$$

Otras representaciones de la función de densidad condicional son las siguientes

$$f_{Y|X}(y \mid x) = f_{Y|X}(y \mid X = x) = f(y \mid X = x) = f(y \mid x).$$

El subíndice es requerido para evitar cualquier confusión de notación. Tanto la función de densidad incondicional (marginal) $f(y)$, como su contraparte condicional $f(y \mid x)$, son predictores de la variable aleatoria Y. La primera es útil al ignorar el valor de la variable aleatoria X. En cambio, la segunda es relevante sólo si se sabe que $\{X = x\}$.

De manera análoga, la función de densidad condicional de la variable aleatoria X, dado el evento $\{Y = y\}$, es

$$f(x \mid y) = \frac{f(x,y)}{f(y)}; \quad \text{para} \quad x, y \in \mathbf{R}, \quad \text{con} \quad f(y) > 0. \tag{4.17}$$

En otros caso, $f(x \mid y) = 0$. Por (4.16) y (4.17), la función de densidad conjunta tiene dos factorizaciones:

$$f(x,y) = f(x) f(y \mid x) = f(x \mid y) f(y); \quad \text{para} \quad (x,y) \in \mathbf{R}^2. \tag{4.18}$$

Esta fórmula describe dos rutas de simulación del vector aleatorio (X, Y). Primero se simula el valor x de la variable aleatoria X; según su función de densidad marginal $f(x)$. Luego, dado el valor $X = x$, se simula el valor y de la variable aleatoria Y; según su función de densidad condicional $f(y \mid x)$. Viceversa, intercambie el rol de x y y.

La función de *distribución condicional* de Y dado $X = x$ es

$$F(y \mid x) = \int_{-\infty}^{y} f_{Y|X}(v \mid x) \mathrm{d}v.$$

Al fijar $x \in \mathbf{R}$, la expresión $F(y \mid x)$ es una función de distribución respecto de $y \in \mathbf{R}$, pues satisface las tres propiedades del Teorema 2.1.3. Esta función se interpreta como la probabilidad condicional del evento $\{Y \leq y\}$, dado el evento $\{X = x\}$:

$$F(y \mid x) = F(y \mid X = x) = P(Y \leq y \mid X = x),$$

en el sentido de

$$F(y \mid x) = \lim_{h \to 0+} P(Y \leq y \mid x \leq X < x + h).$$

Se recuerda que X es una variable aleatoria continua y $P(X = x) = 0$. La distribución condicional tiene asociados los conceptos de esperanza, varianza, entre otros. La *esperanza condicional* de Y dado $X = x$ es

$$\mathrm{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f(y \mid x) \mathrm{d}y; \quad x \in \mathbf{R}. \tag{4.19}$$

La esperanza condicional es una función real, según la relación $g(x) = \mathrm{E}[Y \mid X = x]$; para $x \in \mathbf{R}$. Sin embargo, la esperanza condicional tiene también rol de variable aleatoria:

$$\mathrm{E}[Y \mid X] = \int_{-\infty}^{\infty} y f(y \mid X) \mathrm{d}y.$$

De hecho, al evaluar $g(x)$ en $x = X$, la esperanza condicional es una variable aleatoria en términos de X:

$$g(X) = \mathrm{E}[Y \mid X] = \int_{-\infty}^{\infty} y f(y \mid X) \mathrm{d}y.$$

Ambas representaciones de la esperanza condicional son equivalentes.

Por otro lado, la *varianza condicional* de Y dado $X = x$ es

$$\mathrm{V}[Y \mid X = x] = \mathrm{E}[(Y - \mathrm{E}[Y \mid X = x])^2 \mid X = x] = \int_{-\infty}^{\infty} (y - \mathrm{E}[Y \mid X = x])^2 f(y \mid x) \mathrm{d}x.$$

Note que

$$\mathrm{V}[Y \mid X = x] = \mathrm{E}[Y^2 \mid X = x] - (\mathrm{E}[Y \mid X = x])^2, \quad x \in \mathbf{R},$$

que equivale a

$$\mathrm{V}[Y \mid X] = \mathrm{E}[Y^2 \mid X] - (\mathrm{E}[Y \mid X])^2.$$

De hecho

$$\begin{aligned}
\mathrm{V}[Y \mid X] &= \mathrm{E}[(Y - \mathrm{E}[Y \mid X])^2 \mid X] \\
&= \mathrm{E}[Y^2 - 2Y \cdot \mathrm{E}[Y \mid X] + (\mathrm{E}[Y \mid X])^2 \mid X] \\
&= \mathrm{E}[Y^2 \mid X] - 2\,\mathrm{E}[Y \mid X] \cdot \mathrm{E}[Y \mid X] + (\mathrm{E}[Y \mid X])^2 \\
&= \mathrm{E}[Y^2 \mid X] - (\mathrm{E}[Y \mid X])^2.
\end{aligned}$$

En la penúltima igualdad se usó (2.93).

4.2. Distribución normal bivariada

Un vector aleatorio (X, Y) tiene distribución *normal bivariada* si su función de densidad conjunta es

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}; \quad x, y \in \mathbf{R}, \tag{4.20}$$

con

$$\mu_1, \mu_2 \in \mathbf{R}, \quad \sigma_1^2, \sigma_2^2 > 0, \quad -1 < \rho < 1.$$

Este modelo se caracteriza por 5 parámetros. Los valores μ_1 y μ_2 representan las medias poblacionales de X y Y; respectivamente. Así mismo, sus correspondientes varianzas son σ_1^2 y σ_2^2. El parámetro ρ es el coeficiente de correlación de ambas variables aleatorias. Note que las curvas de nivel de esta función, obtenidas con la igualdad $f(x, y) = c$; para una constante genérica $c > 0$, equivalen a

$$\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 = c',$$

para cierta constante $c' > 0$. Esta es una elipse de centro (μ_1, μ_2). Los ejes menor y mayor corresponden al mínimo y máximo entre σ_1 y σ_2; respectivamente. El coeficiente de correlación ρ se asocia al ángulo de rotación y excentricidad de la elipse. En la Figura 4.3 se muestra la gráfica de las curvas de nivel de nueve funciones de densidad normal bivariadas. Por renglón, se consideran los casos $\sigma_1 = \sigma_2 = 1$, $\sigma_1 = \sigma_2 = 0.5$ y $\sigma_1 = 1 > 0.5 = \sigma_2$. Los casos $\rho = 0, -0.5, 0.9$ se establecen por columna. La unidad de medida es la misma.

Las funciones de densidad marginales son normales (univariadas): $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. Así

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}; \quad x \in \mathbf{R},$$

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2}; \quad y \in \mathbf{R}.$$

Ahora se deducirá la forma de dichas funciones de densidad marginales. Por simplicidad, considere el caso estándar:

$$\mu_1 = \mu_2 = 0 \quad \text{y} \quad \sigma_1^2 = \sigma_2^2 = 1. \tag{4.21}$$

Por (4.10), se tiene

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{x^2 - 2\rho xy + y^2\}}\mathrm{d}y; \quad \text{para} \quad x \in \mathbf{R}. \tag{4.22}$$

Complete el trinomio cuadrado perfecto respecto de y:

$$x^2 - 2\rho xy + y^2 = y^2 - 2\rho xy + \rho^2 x^2 + (1-\rho^2)x^2 = (y - \rho x)^2 + (1-\rho^2)x^2.$$

Así

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{(y-\rho x)^2 + (1-\rho^2)x^2\}}\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2}\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}; \quad \text{para} \quad x \in \mathbf{R}. \end{aligned}$$

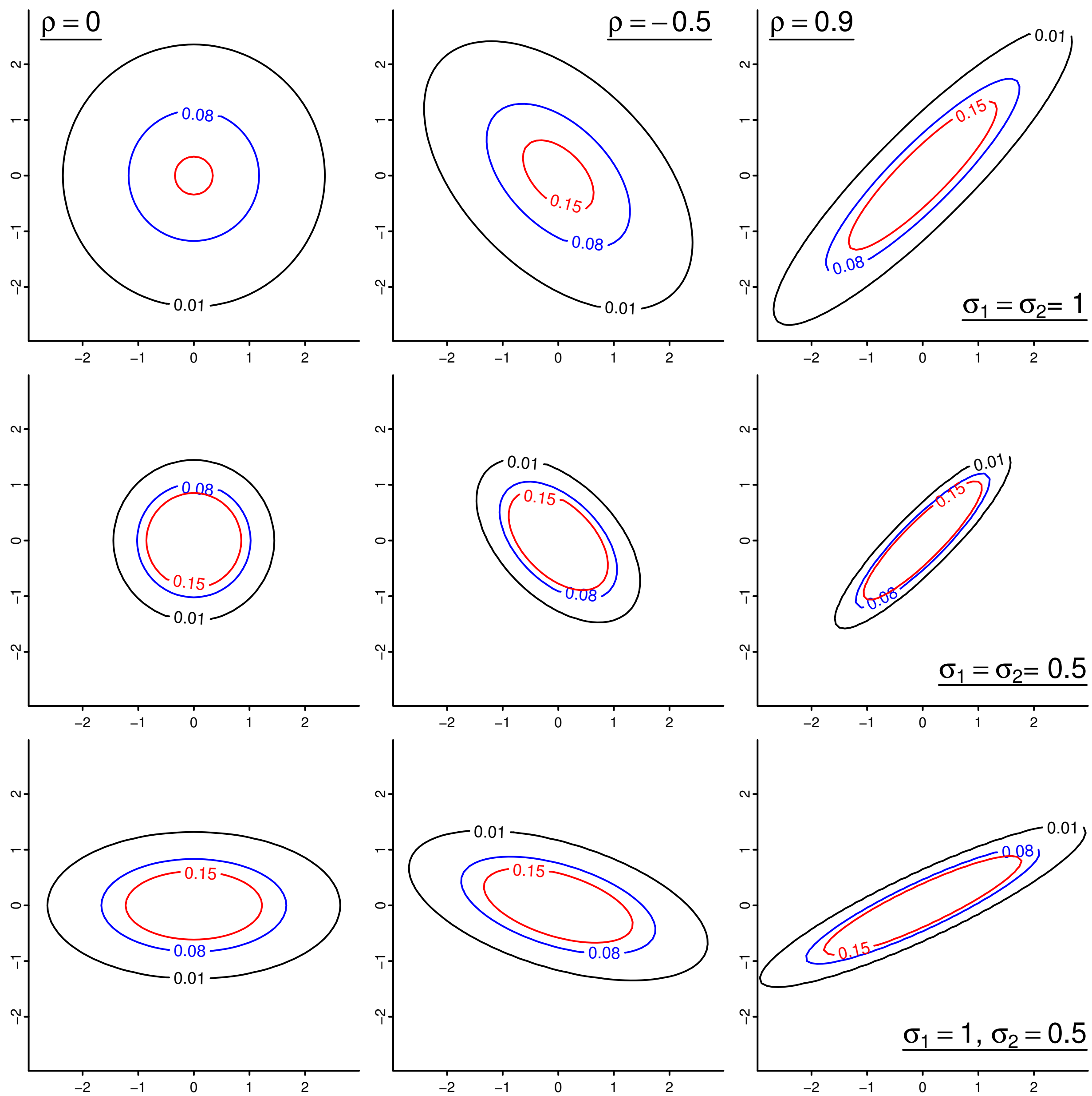


Figura 4.3. Curvas de nivel de la función de densidad normal bivariada. Por renglón se fijan las desviaciones estándar, mientras que el coeficiente de correlación se establece por columna.

El último integrando integra uno, pues corresponde a la función de densidad normal de media ρx y varianza $1-\rho^2$. Por lo tanto, la variable aleatoria X tiene distribución normal estándar. Análogamente, se comprueba que Y y X son iguales en distribución. Por otra parte, al aplicar (4.16), la función de densidad condicional de Y dado $X = x$ es

$$f(y \mid x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} e^{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left\{y-\mu_2-\rho\frac{\sigma_1}{\sigma_2}(x-\mu_1)\right\}^2}; \tag{4.23}$$

para $(x, y) \in \mathbf{R}^2$. En otras palabras

$$[Y \mid X = x] \sim \mathrm{N}\left(\text{media} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \text{varianza} = (1-\rho^2)\sigma_2^2\right). \tag{4.24}$$

La esperanza condicional de Y es una función de dependencia lineal respecto del valor dado $X = x$:

$$\mathrm{E}[Y \mid X = x] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1).$$

Por ejemplo, si $\rho > 0$ y $x > \mu_1$, entonces la media poblacional condicional de Y será mayor que su contraparte incondicional μ_2. Además, disminuye la incertidumbre sobre el valor de la variable aleatoria Y:

$$\mathrm{V}[Y \mid X = x] = (1-\rho^2)\sigma_2^2 < \sigma_2^2 = \mathrm{V}\,Y; \quad \text{para todo} \quad x \in \mathbf{R}.$$

Análogamente, la distribución condicional de X dado $Y = y$ es también normal:

$$[X \mid Y = y] \sim \mathrm{N}\left(\text{media} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \text{varianza} = (1-\rho^2)\sigma_1^2\right). \tag{4.25}$$

En consecuencia

$$\begin{aligned}
\mathrm{E}[X \mid Y = y] &= \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\\
\mathrm{V}[X \mid Y = y] &= (1-\rho^2)\sigma_1^2 \leq \sigma_1^2 = \mathrm{V}\,X.
\end{aligned}$$

Al considerar (4.16), (4.20) y (4.22), se deduce (4.23):

$$\begin{aligned}
f(y \mid x) &= \frac{f(x,y)}{f(x)} \\
&= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}/[2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}]}{\exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\}/[\sqrt{2\pi}\sigma_1]} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} e^{-\frac{1}{2(1-\rho^2)}\left\{[1-(1-\rho^2)]\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} e^{-\frac{1}{2(1-\rho^2)}\left\{\rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right\}^2} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_2} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2\right\}.
\end{aligned}$$

Por último, resta deducir que ρ es el coeficiente de correlación $\rho(X,Y)$. Primero, asuma el caso donde X y Y son variables aleatorias de distribución normal estándar (4.21). Por (4.24), se tiene

$$\mathrm{E}[Y \mid X = x] = \rho x.$$

Al considerar (4.18) y (4.19), resulta

$$\begin{aligned}
\mathrm{E}\,XY &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x)f(y \mid x)\mathrm{d}y\mathrm{d}x \\
&= \int_{-\infty}^{\infty} xf(x)\int_{-\infty}^{\infty} yf(y \mid x)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} xf(x)\,\mathrm{E}[Y \mid X = x]\mathrm{d}x \\
&= \int_{-\infty}^{\infty} xf(x)(\rho x)\mathrm{d}x = \rho\int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x \\
&= \rho\,\mathrm{E}\,X^2 = \rho\,\mathrm{V}\,X \\
&= \rho.
\end{aligned}$$

Por lo tanto

$$\rho(X,Y) = \mathrm{C}(X,Y) = \mathrm{E}\,XY - \mathrm{E}\,X\cdot\mathrm{E}\,Y = \mathrm{E}\,XY = \rho.$$

Segundo, el caso general se deduce con el particular establecido:

$$\rho(X,Y) = \frac{\mathrm{C}(X,Y)}{\sqrt{\mathrm{V}\,X\cdot\mathrm{V}\,Y}} = \frac{\mathrm{E}[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} = \mathrm{E}\left[\frac{X-\mu_1}{\sigma_1}\frac{Y-\mu_2}{\sigma_2}\right] = \rho.$$

Ejemplo 4.2.1. Considere un vector aleatorio (X,Y) de distribución normal en el plano $\mathbf{R}^2$. Compruebe

$$\rho(X^2, Y^2) = \rho^2.$$

Se desarrollará el caso de distribuciones marginales normal estándar: $X \sim \mathrm{N}(0,1)$ y $Y \sim \mathrm{N}(0,1)$. Por (4.20), la función de densidad conjunta de (X,Y) es

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(x^2-2\rho xy+y^2\right)}; \quad (x,y) \in \mathbf{R}^2.$$

Recuerde que las variables aleatorias X y Y son independientes si y sólo si, $\rho = 0$. Por (4.23) y (4.24), la distribución de la variable aleatoria Y dado $X = x$ es normal con

$$\text{media} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) = \rho x \quad \text{y} \quad \text{varianza} = (1-\rho^2)\sigma_2^2 = 1-\rho^2.$$

Así, la función de densidad condicional de Y dado $X = x$ es

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2}; \quad (x,y) \in \mathbf{R}^2.$$

Entonces

$$\begin{aligned}
\mathrm{V}\,X^2 = \mathrm{V}\,Y^2 &= \mathrm{E}(X^2)^2 - (\mathrm{E}\,X^2)^2 = \mathrm{E}\,X^4 - 1 = 3 - 1 = 2,\\
\mathrm{C}(X^2,Y^2) &= \mathrm{E}\,X^2Y^2 - \mathrm{E}\,X^2 \cdot \mathrm{E}\,Y^2 = \mathrm{E}\,X^2Y^2 - 1,
\end{aligned}$$

con

$$\begin{aligned}
\mathrm{E}\,X^2Y^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2y^2 f(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2y^2 f(x)f(y\mid x)\mathrm{d}y\mathrm{d}x\\
&= \int_{-\infty}^{\infty} x^2 f(x) \int_{-\infty}^{\infty} y^2 f(y\mid x)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{E}[Y^2 \mid X = x]\mathrm{d}x\\
&= \int_{-\infty}^{\infty} x^2 f(x)[\mathrm{V}[Y \mid X = x] + \{\mathrm{E}[Y \mid X = x]\}^2]\mathrm{d}x\\
&= \int_{-\infty}^{\infty} x^2 f(x)[(1-\rho^2) + (\rho x)^2]\mathrm{d}x\\
&= (1-\rho^2)\int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x + \rho^2 \int_{-\infty}^{\infty} x^4 f(x)\mathrm{d}x\\
&= (1-\rho^2)\,\mathrm{E}\,X^2 + \rho^2\,\mathrm{E}\,X^4 = 1 - \rho^2 + 3\rho^2\\
&= 1 + 2\rho^2.
\end{aligned}$$

Por lo tanto

$$\rho(X^2,Y^2) = \frac{\mathrm{C}(X^2,Y^2)}{\sqrt{\mathrm{V}\,X^2 \cdot \mathrm{V}\,Y^2}} = \frac{1+2\rho^2-1}{\sqrt{2\cdot 2}} = \rho^2 < \rho.$$

Distribución normal bivariada: enfoque matricial.* Sea (X_1, X_2) un vector aleatorio de distribución normal bivariada. Su función de densidad (4.20) se puede escribir como

$$f(x) = f(x_1,x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2}; \quad x = (x_1,x_2) \in \mathbf{R}^2, \tag{4.26}$$

donde el *vector de medias* es

$$\mu = \mathrm{E}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mathrm{E}\,X_1 \\ \mathrm{E}\,X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

mientras que la *matriz de covarianzas* es

$$\Sigma = \mathrm{E}\left[\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}'\right] = \begin{pmatrix} \mathrm{V}\,X_1 & \mathrm{C}(X_1,X_2) \\ \mathrm{C}(X_1,X_2) & \mathrm{V}\,X_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Como la matriz Σ es positiva definida, su determinante es positivo

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = (1-\rho^2)\sigma_1^2\sigma_2^2 > 0$$

y

$$|\Sigma|^{1/2} = \sigma_1\sigma_2\sqrt{1-\rho^2}.$$

Véase Sección F del Apéndice. Así mismo, la matriz Σ es invertible cuya inversa es también simétrica. Por (25) y la forma cuadrática (24) del Apéndice, se tiene

$$\Sigma^{-1} = \frac{1}{(1-\rho^2)\,\sigma_1^2\sigma_2^2}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

y

$$\begin{aligned}(x-\mu)'\Sigma^{-1}(x-\mu) &= \frac{\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2}{(1-\rho^2)\sigma_1^2\sigma_2^2} \\ &= \frac{1}{1-\rho^2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_2}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right],\end{aligned}$$

para todo $x = (x_1, x_2) \in \mathbf{R}^2$. Se concluye que (4.20) y (4.26) coinciden.

Ejemplo 4.2.2. Sea la función real

$$f(x,y) = ke^{-2x-3y}; \quad x, y > 0,$$

para cierta constante $k > 0$. Encuentre la constante k tal que $f(x,y)$ sea una función de densidad bivariada. Obtenga las funciones de densidad marginales y condicionales.

Solución. Por (4.4) y (4.5), la función $f(x,y)$ es una función de densidad si $f(x,y) \geq 0$; para todo $(x,y) \in \mathbf{R}^2$ e $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = 1$. La primera propiedad se cumple claramente. Además, el conjunto de valores posibles del vector aleatorio (X,Y) es el primer cuadrante del plano. Por lo que

$$\begin{aligned}1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = \int_0^{\infty}\int_0^{\infty} ke^{-2x-3y}\mathrm{d}y\mathrm{d}x \\ &= k\int_0^{\infty} e^{-2x}\int_0^{\infty} e^{-3y}\mathrm{d}y\mathrm{d}x = k\int_0^{\infty} e^{-2x}\mathrm{d}x\int_0^{\infty} e^{-3y}\mathrm{d}y \\ &= \frac{k}{2}\int_0^{\infty} 2e^{-2x}\mathrm{d}x\frac{1}{3}\int_0^{\infty} 3e^{-3y}\mathrm{d}y = \frac{k}{6}.\end{aligned}$$

Por lo que k debe ser 6. Por otro lado, la forma de la función de densidad conjunta $f(x,y)$ sugiere que X y Y sean independientes, puesto que

$$f(x,y) = 6e^{-2x-3y} = 6e^{-2x} \cdot e^{-3y} = 2e^{-2x} \cdot 3e^{-3y}; \quad x,y > 0.$$

Este es un producto de dos funciones de densidad exponencial de tasas $\lambda_1 = 2$ y $\lambda_2 = 3$. Por lo tanto, $X \sim \exp(\lambda = 2)$ y $Y \sim \exp(\lambda = 3)$. De hecho, la función de densidad marginal de X se deduce como sigue.

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = \int_{0}^{\infty} 6e^{-2x} \cdot e^{-3y}\mathrm{d}y \\ &= 2e^{-2x}\int_0^\infty 3e^{-3y}\mathrm{d}y = 2e^{-2x}; \quad x > 0. \end{aligned}$$

De la misma manera, se ratifica que Y tiene una distribución exponencial:

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x = \int_{0}^{\infty} 6e^{-2x} \cdot e^{-3y}\mathrm{d}x \\ &= 3e^{-3y}\int_0^\infty 2e^{-2x}\mathrm{d}x = 3e^{-3y}; \quad y > 0. \end{aligned}$$

Como

$$f(x,y) = 6e^{-2x-3y} = 2e^{-2x} \cdot 3e^{-2y} = f(x)f(y); \quad \text{para} \quad x,y > 0,$$

entonces X y Y son variables aleatorias independientes. Por lo tanto, la función de densidad condicional de Y dado X coincide con su contraparte marginal:

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{f(x)f(y)}{f(x)} = f(y) = 3e^{-3y}; \quad y > 0.$$

Así mismo la función de densidad condicional de X dado Y es su contraparte incondicional:

$$f(x \mid y) = \frac{f(x,y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x) = 2e^{-2x}; \quad x > 0.$$

Ejemplo 4.2.3. Sea la función real

$$f(x,y) = ke^{-2x-3y}; \quad 0 < x < y,$$

para cierta constante $k > 0$. Encuentre la constante k tal que $f(x,y)$ sea una función de densidad bivariada. Obtenga además las densidades marginales y condicionales.
Solución. Recuerde que $f(x,y)$ es una función de densidad si es no negativa e integra uno: (4.4)-(4.5). La primera condición es clara. El conjunto de vectores posibles de (X,Y) es la región de integración formada por el triángulo entre las curvas $y = 0$ y $y = x$, del primer cuadrante:

$$\{(x,y) \in \mathbf{R}^2 : 0 < x < y\} \subset \mathbf{R}^2.$$

Por lo que

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = \int_{0}^{\infty}\int_{x}^{\infty} ke^{-2x-3y}\mathrm{d}y\mathrm{d}x \\
&= k\int_{0}^{\infty} e^{-2x}\int_{x}^{\infty} e^{-3y}\mathrm{d}y\mathrm{d}x = \frac{k}{3}\int_{0}^{\infty} e^{-2x}\int_{x}^{\infty} 3e^{-3y}\mathrm{d}y\mathrm{d}x \\
&= \frac{k}{3}\int_{0}^{\infty} e^{-2x}\left[-e^{-u}\right]_{3x}^{\infty}\mathrm{d}x = \frac{k}{3}\int_{0}^{\infty} e^{-2x}e^{-3x}\mathrm{d}x \\
&= \frac{k}{3}\int_{0}^{\infty} e^{-5x}\mathrm{d}x = \frac{k}{3\times 5}\int_{0}^{\infty} 5e^{-5x}\mathrm{d}x \\
&= \frac{k}{15}.
\end{aligned}$$

El valor de k debe ser 15. Por lo que la función de densidad conjunta queda como

$$f(x,y) = 15e^{-2x-3y}; \quad 0 < x < y.$$

Por el Teorema de Tonelli Fubini, esta función de densidad se puede obtener también al intercambiar los límites de integración. En este caso, la región de integración se escribe como

$$\begin{aligned}
\{(x,y)\in\mathbf{R}^2 : 0 < x < y\} &= \{(x,y)\in\mathbf{R}^2 : x > 0, y > x\} \\
&= \{(x,y)\in\mathbf{R}^2 : y > 0, 0 < x < y\}
\end{aligned}$$

y

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{0}^{\infty}\int_{0}^{y} ke^{-2x-3y}\mathrm{d}x\mathrm{d}y = \frac{k}{2}\int_{0}^{\infty} e^{-3y}\int_{0}^{y} 2e^{-2x}\mathrm{d}x\mathrm{d}y \\
&= \frac{k}{2}\int_{0}^{\infty} e^{-3y}\left[-e^{-u}\right]_{0}^{2y}\mathrm{d}y = \frac{k}{2}\int_{0}^{\infty} e^{-3y}\left[1-e^{-2y}\right]\mathrm{d}y \\
&= \frac{k}{2}\int_{0}^{\infty}(e^{-3y}-e^{-5y})\mathrm{d}y = \frac{k}{2}\left[\frac{1}{3}\int_{0}^{\infty} 3e^{-3y}\mathrm{d}y - \frac{1}{5}\int_{0}^{\infty} 5e^{-5y}\mathrm{d}y\right] \\
&= \frac{k}{2}\left[\frac{1}{3}-\frac{1}{5}\right] = \frac{k}{15}.
\end{aligned}$$

Note que X y Y son variables aleatorias dependientes, pues $X < Y$, con probabilidad uno. Compare con el ejemplo anterior. Ahora se obtendrán las funciones de densidad marginales. El conjunto de valores posibles de la variable aleatoria X es

$$\mathcal{X} = \{x\in\mathbf{R} : x > 0\} = (0,\infty).$$

Así, para $x > 0$ fijo, se tiene

$$\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y = \int_{x}^{\infty} 15e^{-2x-3y}\mathrm{d}y \\
&= 15e^{-2x}\int_{x}^{\infty} e^{-3y}\mathrm{d}y = \frac{15}{3}e^{-2x}\int_{x}^{\infty} e^{-3y}(3\mathrm{d}y) \\
&= 5e^{-2x}\left\{-e^{-u}\right\}_{3x}^{\infty} = 5e^{-2x-3x} \\
&= 5e^{-5x}.
\end{aligned}$$

Por lo tanto, X tiene una distribución marginal exponencial de tasa $\lambda_1 = 5$. De la misma manera, el conjunto de valores posibles de Y coincide con el de X:

$$\mathcal{Y} = \{y \in \mathbf{R} : y > 0\} = (0, \infty).$$

Por lo que

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x = \int_0^y 15e^{-2x-3y}\mathrm{d}x \\ &= 15e^{-3y}\int_0^y e^{-2x}\mathrm{d}x = \frac{15}{2}e^{-3y}\int_0^y e^{-2x}(2\mathrm{d}x) \\ &= 7.5e^{-3y}\left\{-e^{-u}\right\}_0^{2y} = 7.5e^{-3y}\left(1-e^{-2y}\right); \quad y > 0. \end{aligned}$$

Se puede confirmar que $f(y)$ es una función de densidad:

$$\begin{aligned} \int_{-\infty}^{\infty} f(y)\mathrm{d}y &= \int_0^{\infty} 7.5e^{-3y}\left(1-e^{-2y}\right)\mathrm{d}y = 7.5\int_0^{\infty}\left(e^{-3y}-e^{-5y}\right)\mathrm{d}y \\ &= 7.5\int_0^{\infty}\left(\frac{1}{3}3e^{-3y} - \frac{1}{5}5e^{-5y}\right)\mathrm{d}y = 7.5\left(\frac{1}{3}-\frac{1}{5}\right) \\ &= 7.5\frac{5-3}{15} = 1. \end{aligned}$$

Por último, se obtendrán las densidades condicionales. Dado $X = x > 0$, la variable aleatoria Y sólo toma valores mayores que x:

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{15e^{-2x-3y}}{5e^{-5x}} = 3e^{3x-3y} = 3e^{-3(y-x)}; \quad 0 < x < y.$$

Por la forma de esta función, ¿que distribución de probabilidad es? Por otro lado, dado $Y = y > 0$, la variable aleatoria X sólo toma valores en $(0, y)$:

$$f(x \mid y) = \frac{f(x,y)}{f(y)} = \frac{15e^{-2x-3y}}{7.5e^{-3y}\left(1-e^{-2y}\right)} = \frac{2e^{-2x}}{1-e^{-2y}}; \quad 0 < x < y.$$

Esta es la distribución de probabilidad exponencial de tasa $\lambda = 2$, truncada a la derecha en $y > 0$. La forma de ambas funciones de densidad condicionales exhiben que X y Y son variables aleatorias dependientes. Por otro lado, se ha afirmado que las funciones de densidad condicionales son también funciones de densidad: son no negativas e integran uno. De hecho

$$\int_{-\infty}^{\infty} f(y \mid x)\mathrm{d}y = \int_x^{\infty} 3e^{-3(y-x)}\mathrm{d}y = \int_0^{\infty} 3e^{-3u}\mathrm{d}u = 1$$

y

$$\int_{-\infty}^{\infty} f(x \mid y)\mathrm{d}x = \int_0^y \frac{2e^{-2x}}{1-e^{-2y}}\mathrm{d}x = \frac{1}{1-e^{-2y}}\int_0^y e^{-2x}(2\mathrm{d}x) = \frac{\left\{-e^{-u}\right\}_0^{2y}}{1-e^{-2y}} = \frac{1-e^{-2y}}{1-e^{-2y}} = 1.$$

4.3. Transformación de vectores aleatorios

Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio continuo de dimensión $n \geq 1$, con conjunto de vectores posibles $\mathcal{X} \subset \mathbf{R}^n$ y función de densidad

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n); \quad x = (x_1, \ldots, x_n) \in \mathcal{X}.$$

Considere la transformación

$$Y = (Y_1, \ldots, Y_n) = g(X_1, \ldots, X_n) = (g_1(X_1, \ldots, X_n), \ldots, g_n(X_1, \ldots, X_n)),$$

donde $g = (g_1, \ldots, g_n)$ es una transformación diferenciable, uno a uno, que mapea $\mathcal{X} \subset \mathbf{R}^n$ sobre $\mathcal{Y} \subset \mathbf{R}^n$:

$$g(x) = (g_1(x), \ldots, g_n(x)) = (g_1(x_1, \ldots, x_n), \ldots, g_n(x_1, \ldots, x_n));$$

para $x = (x_1, \ldots, x_n) \in \mathcal{X}$. Denote como $h(y)$ su función inversa:

$$h(y) = (h_1(y), \ldots, h_n(y)) = (h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n));$$

para $y = (y_1, \ldots, y_n) \in \mathcal{Y}$. Su *jacobiano* es

$$J = \frac{\partial h(y)}{\partial y} = \begin{bmatrix} \frac{\partial}{\partial y_1} h_1(y_1, \ldots, y_n) & \frac{\partial}{\partial y_2} h_1(y_1, \ldots, y_n) & \cdots & \frac{\partial}{\partial y_n} h_1(y_1, \ldots, y_n) \\ \frac{\partial}{\partial y_1} h_2(y_1, \ldots, y_n) & \frac{\partial}{\partial y_2} h_2(y_1, \ldots, y_n) & \cdots & \frac{\partial}{\partial y_n} h_2(y_1, \ldots, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial y_1} h_n(y_1, \ldots, y_n) & \frac{\partial}{\partial y_2} h_n(y_1, \ldots, y_n) & & \frac{\partial}{\partial y_n} h_n(y_1, \ldots, y_n) \end{bmatrix}.$$

Entonces, la función de densidad conjunta de Y es

$$f(y) = |\det J| f(h(y)); \quad \text{para} \quad y \in \mathcal{Y}, \tag{4.27}$$

donde $|\det J| > 0$ denota el valor absoluto del determinante del jacobiano J. Informalmente, este resultado se escribe como

$$f(y_1, \ldots, y_n) = |\det J| f(x_1, \ldots, x_n),$$

con

$$x_1 = h_1(y_1, \ldots, y_n), \ldots, x_n = h_n(y_1, \ldots, y_n)$$

y

$$J = \frac{\partial x}{\partial y} = \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \begin{bmatrix} \frac{\partial}{\partial y_1} x_1 & \frac{\partial}{\partial y_2} x_1 & \cdots & \frac{\partial}{\partial y_n} x_1 \\ \frac{\partial}{\partial y_1} x_2 & \frac{\partial}{\partial y_2} x_2 & \cdots & \frac{\partial}{\partial y_n} x_2 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial y_1} x_n & \frac{\partial}{\partial y_2} x_n & & \frac{\partial}{\partial y_n} x_n \end{bmatrix}.$$

La fórmula (4.27) es la versión multivariada de (3.60).

Ejemplo 4.3.1. Sean X_1, X_2 dos variables aleatorias independientes de distribución idéntica gamma, de parámetros $\alpha > 0$ y $\beta = 1$. La función de densidad conjunta es

$$\begin{aligned} f(x_1, x_2) = f(x_1)f(x_2) &= \frac{1}{\Gamma(\alpha)} x_1^{\alpha-1} e^{-x_1} \frac{1}{\Gamma(\alpha)} x_2^{\alpha-1} e^{-x_2} \\ &= \frac{1}{\Gamma^2(\alpha)} (x_1 x_2)^{\alpha-1} e^{-(x_1+x_2)}; \quad x_1, x_2 > 0. \end{aligned}$$

Defina

$$Y_1 = \frac{X_1}{X_1 + X_2} \quad \text{y} \quad Y_2 = X_1 + X_2.$$

Luego

$$\begin{aligned} \mathcal{X} &= \{(x_1, x_2) \in \mathbf{R}^2 : x_1, x_2 > 0\}, \\ \mathcal{Y} &= \{(y_1, y_2) \in \mathbf{R}^2 : 0 < y_1 < 1, y_2 > 0\}. \end{aligned}$$

y

$$\begin{aligned} y_1 &= \frac{x_1}{x_1 + x_2}, \quad y_2 = x_1 + x_2, \\ x_1 &= y_1 y_2, \quad x_2 = (1 - y_1) y_2 \end{aligned}$$

Así

$$J = \frac{\partial h(y_1, y_2)}{\partial(y_1, y_2)} = \begin{bmatrix} \frac{\partial}{\partial y_1}\{y_1 y_2\} & \frac{\partial}{\partial y_2}\{y_1 y_2\} \\ \frac{\partial}{\partial y_1}\{(1 - y_1) y_2\} & \frac{\partial}{\partial y_2}\{(1 - y_1) y_2\} \end{bmatrix} = \begin{bmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{bmatrix},$$

$$|\det J| = |y_2(1 - y_1) + y_2 y_1| = y_2 > 0$$

y

$$\begin{aligned} f(y_1, y_2) &= |\det J| f(x_1, x_2) = y_2 f(x_1) f(x_2) \\ &= y_2 f_{X_1}(y_1 y_2) f_{X_2}((1 - y_1) y_2) \\ &= y_2 \frac{1}{\Gamma(\alpha)} (y_1 y_2)^{\alpha-1} e^{-y_1 y_2} \times \frac{1}{\Gamma(\alpha)} [(1 - y_1) y_2]^{\alpha-1} e^{-(1-y_1) y_2} \\ &= \frac{1}{\Gamma^2(\alpha)} [y_1(1 - y_1)]^{\alpha-1} y_2^{2(\alpha-1)+1} e^{-y_1 y_2 - y_2 + y_1 y_2} \\ &= \frac{1}{\Gamma^2(\alpha)} [y_1(1 - y_1)]^{\alpha-1} y_2^{2\alpha-1} e^{-y_2}; \quad 0 < y_1 < 1, \quad y_2 > 0. \end{aligned}$$

Note la independencia de las variables aleatorias Y_1 y Y_2, pues su función de densidad conjunta es el producto de dos funciones reales

$$f(y_1, y_2) = g_1(y_1) g_2(y_2),$$

donde

$$g_1(y_1) = \frac{1}{\Gamma^2(\alpha)} [y_1(1 - y_1)]^{\alpha-1} \quad \text{y} \quad g_2(y_2) = y_2^{2\alpha-1} e^{-y_2}; \quad 0 < y_1 < 1, \quad y_2 > 0.$$

La forma de la primera función se asocia a la función de densidad beta(forma$_1 = \alpha$, forma$_2 = \alpha$), mientras que la segunda corresponde a una gamma(forma $= 2\alpha$, tasa $= 1$). Así

$$f(y_1) = \frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)} y_1(1-y_1) = \frac{y_1^{\alpha-1}(1-y_1)^{\alpha-1}}{\mathrm{B}(\alpha,\alpha)}; \quad 0 < y_1 < 1$$

y

$$f(y_2) = \frac{1}{\Gamma(2\alpha)} y_2^{2\alpha-1} e^{-y_2}; \quad y_2 > 0$$

Aquí se usó (3.54) y (3.55).

Ejemplo 4.3.2. Sea $X = (X_1, \ldots, X_n)$ un vector aleatorio de distribución normal de dimensión n, con de vector de medias $\mu \in \mathbf{R}^n$ y matriz de covarianza Σ de $n \times n$. Defina el vector aleatorio $Y = AX + b$, donde A es una matriz invertible de $n \times n$ y $b \in \mathbf{R}^n$ es un vector constante. Entonces, por la fórmula del teorema de transformación (4.27), Y tiene distribución normal multivariada de media $A\mu + b$ y matriz de covarianza $A\Sigma A'$. Considere

$$\begin{aligned} g(x) &= Ax + b, \\ y &= Ax + b, \\ y - b &= Ax, \\ A^{-1}(y-b) &= x, \\ h(y) &= A^{-1}(y-b), \end{aligned}$$

La función inversa $h(y)$ tiene Jacobiano

$$J = \frac{\partial}{\partial y} h(y) = \frac{\partial}{\partial y}\left[A^{-1}(y-b)\right] = A^{-1} I_n = A^{-1}.$$

Aquí I_n denota la matriz identidad en $\mathbf{R}^n$. Por lo que

$$\begin{aligned} f(y) = |\det J| f_X(h(y)) &= |A^{-1}| \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-(h(y)-\mu)'\Sigma^{-1}(h(y)-\mu)/2} \\ &= \frac{1}{(2\pi)^{n/2}|A||\Sigma|^{1/2}} e^{-\left(A^{-1}(y-b)-\mu\right)'\Sigma^{-1}\left(A^{-1}(y-b)-\mu\right)/2}, \end{aligned}$$

con

$$\begin{aligned} \left(A^{-1}(y-b)-\mu\right)' \Sigma^{-1} \left(A^{-1}(y-b)-\mu\right) &= \left(A^{-1}(y-b-A\mu)\right)' \Sigma^{-1} \left(A^{-1}(y-b-A\mu)\right) \\ &= (y-b-A\mu)' \left(A^{-1}\right)' \Sigma^{-1} A^{-1}(y-b-A\mu) \\ &= (y-b-A\mu)'(A')^{-1}\Sigma^{-1}A^{-1}(y-b-A\mu) \\ &= (y-A\mu-b)'(A\Sigma A')^{-1}(y-A\mu-b) \end{aligned}$$

y

$$|A\Sigma A'| = |A||\Sigma||A'| = |A|^2|\Sigma| = \left(|A||\Sigma|^{1/2}\right)^2.$$

Entonces

$$f(y) = \frac{1}{(2\pi)^{n/2}|A\Sigma A'|^{1/2}} e^{-(y-A\mu-b)'(A\Sigma A')^{-1}(y-A\mu-b)/2}; \quad y \in \mathbf{R}^n.$$

Este resultado se generaliza como sigue. Si $X \sim \mathrm{N}(\mu, \Sigma)$, entonces

$$Y \sim \mathrm{N}(A\mu + b, A\Sigma A'),$$

donde A es una matriz de $m \times n$ y b es un vector constante dimensión m, tal que $1 \leq m \leq n$ y $A\Sigma A'$ sea invertible.

Ejemplo 4.3.3. Sean X_1, X_2, X_3 tres variables aleatorias independientes de distribución uniforme continua en el intervalo $(0,1)$. Se obtendrán las funciones de densidad de las medias muestrales $\bar{X}_2 = (X_1 + X_2)/2$ y $\bar{X}_3 = (X_1 + X_2 + X_3)/3$. Sea $Y_2 = X_1 + X_2$. Entonces

$$\begin{aligned}
&x_1 = x_1, \quad y_2 = x_1 + x_2; \quad \text{con} \quad 0 < x_1, x_2 < 1, \\
&x_1 = x_1, \quad x_2 = y_2 - x_1; \quad \text{con} \quad 0 < x_1 < 1, \quad 0 < y_2 - x_1 < 1.
\end{aligned}$$

Esta región de integración equivale a cualesquiera de las siguientes

$$\begin{aligned}
&0 < x_1 < 1, \quad x_1 < y_2 < x_1 + 1, \\
&0 < y_2 < 2, \quad y_2 - 1 < x_1 < y_2, \quad 0 < x_1 < 1, \\
&0 < y_2 < 2, \quad 0 \vee (y_2 - 1) < x_1 < (y_2 \wedge 1).
\end{aligned} \tag{4.28}$$

El jacobiano satisface

$$J = \frac{\partial(x_1, x_2)}{\partial(x_1, y_2)} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \quad \text{y} \quad |\det J| = 1.$$

Así que, la función de densidad conjunta de (X_1, Y_2) es

$$f(x_1, y_2) = f(x_1, x_2)|\det J| = f(x_1)f(x_2) = 1.$$

Se aclara que X_1 y Y_2 son variables aleatorias dependientes, por la forma de la región de integración (4.28). La función de densidad marginal de Y_2 es

$$\begin{aligned}
f(y_2) &= \int_{-\infty}^{\infty} f(x_1, y_2)\mathrm{d}x_1 = \int_{0\vee(y_2-1)}^{1\wedge y_2} \mathrm{d}x_1 \\
&= [1 \wedge y_2] - [0 \vee (y_2 - 1)] = \begin{cases} y_2 & 0 < y_2 \leq 1 \\ 2 - y_2 & 1 \leq y_2 < 2 \end{cases} \\
&= y_2 \wedge (2 - y_2); \quad \text{para} \quad 0 < y_2 < 2.
\end{aligned} \tag{4.29}$$

Esta función es simétrica respecto de $y_2 = 1$, así como integra uno:

$$\begin{aligned}
\int_{-\infty}^{\infty} f(y_2)\mathrm{d}y_2 &= \int_0^2 [y \vee (2 - y)]\mathrm{d}y = \int_0^1 y\mathrm{d}y + \int_1^2 (2 - y)\mathrm{d}y \\
&= \left[\frac{y^2}{2}\right]_0^1 + \left[2y - \frac{y^2}{2}\right]_1^2 = 1.
\end{aligned}$$

Por su forma, se dice que la variable aleatoria Y_2 tiene distribución *triangular* en el intervalo $(0,2)$. En este sentido, por la fórmula del teorema de transformación (3.60) y (4.29), la función de densidad de $\bar{X}_2 = Y_2/2$, tiene distribución triangular en el intervalo $(0,1)$:

$$f(x) = 2 \times [(2x) \wedge (2-2x)] = 4[x \wedge (1-x)]; \quad \text{para} \quad 0 < x < 1.$$

Por otro lado, la función de densidad de $Y_3 = X_1 + X_2 + X_3 = Y_2 + X_3$, se obtendrá vía la función de densidad conjunta de (Y_2, Y_3). Considere

$$\begin{aligned} y_2 &= y_2, \quad y_3 = y_2 + x_3; \quad \text{con} \quad 0 < y_2 < 2, \quad 0 < x_3 < 1, \\ y_2 &= y_2, \quad x_3 = y_3 - y_2; \quad \text{con} \quad 0 < y_2 < 2, \quad 0 < y_3 - y_2 < 1. \end{aligned}$$

El conjunto de vectores posibles de (Y_2, Y_3) es cualesquiera de las siguientes formas equivalentes

$$\begin{aligned} &0 < y_2 < 2, \quad y_2 < y_3 < y_2 + 1, \\ &0 < y_3 < 3, \quad y_3 - 1 < y_2 < y_3, \quad 0 < y_2 < 2, \\ &0 < y_3 < 3, \quad 0 \vee (y_3 - 1) < y_2 < (y_3 \wedge 2). \end{aligned} \tag{4.30}$$

Como

$$J = \frac{\partial(y_2, x_3)}{\partial(y_2, y_3)} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \quad \text{y} \quad |\det J| = 1,$$

entonces la función de densidad de (Y_2, Y_3) es

$$\begin{aligned} f(y_2, y_3) &= f(y_2, x_3) = f(y_2) f(x_3) \\ &= f_{Y_2}(y_2) f_{X_3}(y_3 - y_2) = [y_2 \wedge (2 - y_2)] \times 1 \\ &= y_2 \wedge (2 - y_2). \end{aligned}$$

De nuevo, por la forma de la región de integración (4.30), se aclara que Y_2 y Y_3 son variables aleatorias dependientes. La función de densidad marginal de Y_3 es

$$f(y_3) = \int_{-\infty}^{\infty} f(y_2, y_3) \mathrm{d}y_2 = \int_{0 \vee (y_3 - 1)}^{y_3 \wedge 2} [y_2 \wedge (2 - y_2)] \mathrm{d}y_2.$$

Considere tres casos. Si $0 < y_3 < 1$, entonces $0 < y_2 < 1$ y

$$f(y_3) = \int_0^{y_3} y_2 \mathrm{d}y_2 = \frac{y_3^2}{2}.$$

De la misma manera, si $2 < y_3 < 3$, entonces $1 < y_3 - 1 < y_2 < 2$ y

$$f(y_3) = \int_{y_3 - 1}^{2} (2 - y_2) \mathrm{d}y_2 = -\int_{3 - y_3}^{0} u \mathrm{d}u = \int_0^{3 - y_3} u \mathrm{d}u = \frac{(3 - y_3)^2}{2}.$$

El caso $1 < y_3 < 2$, implica $0 < y_3 - 1 < y_2 < y_3$ (> 1) y

$$\begin{aligned} f(y_3) &= \int_{y_3 - 1}^{y_3} [y_2 \wedge (2 - y_2)] \mathrm{d}y_2 = \int_{y_3 - 1}^{1} y_2 \mathrm{d}y_2 + \int_1^{y_3} (2 - y_2) \mathrm{d}y_2 \\ &= \frac{1 - (y_3 - 1)^2}{2} - \int_1^{2 - y_3} u \mathrm{d}u = \frac{1 - (y_3 - 1)^2}{2} - \frac{(2 - y_3)^2 - 1}{2} \\ &= -y_3^2 + 3y_3 - \frac{3}{2} = \frac{3}{4} - \left(y_3 - \frac{3}{2}\right)^2. \end{aligned}$$

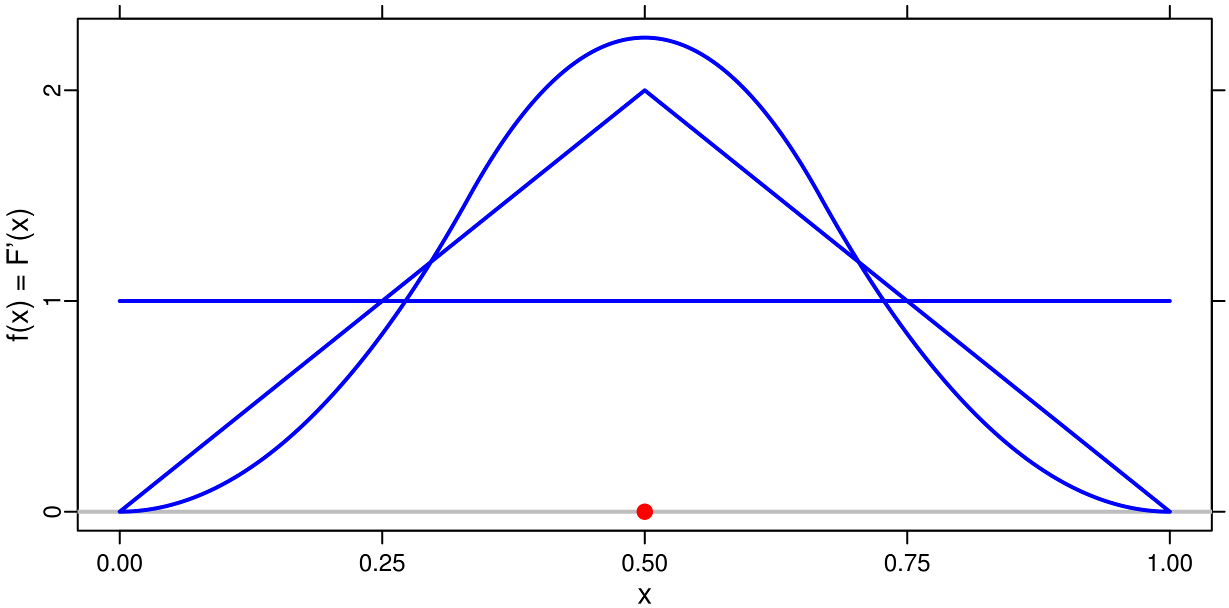


Figura 4.4. Gráfica de la función de densidad de la media muestral $\bar{X}_n$, de una población uniforme continua en $(0,1)$; para $n = 1, 2, 3$.

Se concluye que la función de densidad de Y_3 es la función continua

$$f(y_3) = \begin{cases} y_3^2/2 & 0 < y_3 \leq 1 \\ 3/4 - (y_3 - 3/2)^2 & 1 \leq y_3 \leq 2 \\ (3 - y_3)^2/2 & 2 \leq y_3 < 3. \end{cases}$$

Por la fórmula del teorema de transformación (3.60), la función de densidad de $\bar{X}_3 = Y_3/3$ es

$$\begin{aligned} f(x) &= 3 \begin{cases} (3x)^2/2 & 0 < x \leq 1/3 \\ 3/4 - (3x - 3/2)^2 & 1/3 \leq x \leq 2/3 \\ (3 - 3x)^2/2 & 2/3 \leq x < 1 \end{cases} \\ &= 3 \begin{cases} (9/2)x^2 & 0 < x \leq 1/3 \\ 3/4 - 9(x - 1/2)^2 & 1/3 \leq x \leq 2/3 \\ (9/2)(1 - x)^2 & 2/3 \leq x < 1. \end{cases} \end{aligned}$$

La Figura 4.4 muestra la gráfica de la función de densidad de $\bar{X}_n$; para $n = 1, 2, 3$. Note que la distribución de la media muestral es simétrica respecto de la media poblacional $\mu = 0.5$ mientras que su dispersión disminuye al crecer el tamaño de muestra n.

4.4. Distribución de la suma y cociente de variables aleatorias

En esta sección se deducirá la función de densidad de la suma y cociente de dos variables aleatoria continuas.

Sean X y Y dos variables aleatorias continuas con función de densidad conjunta $f(x, y)$. Considere la transformación $Z = g(X, Y)$, donde $g(x, y)$ es una función real continua. Su función de distribución es

$$F(z) = P(Z \leq z) = P(g(X, Y) \leq z) = P((X, Y) \in A_z) = \iint_{A_z} f(x, y)\mathrm{d}x\mathrm{d}y, \tag{4.31}$$

donde

$$A_z = \{(x, y) \in \mathbf{R}^2 : g(x, y) \leq z\}; \quad z \in \mathbf{R}.$$

En particular si $Z = X + Y$, entonces $g(x, y) = x + y$ y

$$\begin{aligned} A_z &= \{(x, y) : x + y \leq z\} \\ &= \{(x, y) : -\infty < x < \infty, -\infty < y \leq z - x\} \\ &= \{(x, y) : -\infty < y < \infty, -\infty < x \leq z - y\}; \quad z \in \mathbf{R}. \end{aligned} \tag{4.32}$$

La región A_z es el semiplano bajo la curva $y = z - x$; que se muestra en la Figura 4.5. Así

$$F(z) = P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x, y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x, y)\mathrm{d}x\mathrm{d}y;$$

para $z \in \mathbf{R}$. La función de densidad de Z es

$$\begin{aligned} f(z) = F'(z) &= \frac{\mathrm{d}}{\mathrm{d}z}\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x, y)\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}z}\int_{-\infty}^{z-x} f(x, y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}z}\int_{-\infty}^{z-y} f(x, y)\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\mathrm{d}y. \end{aligned}$$

Si además X y Y son variables aleatorias independientes, entonces $f(x, y) = f(x)f(y)$ y

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\mathrm{d}x = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\mathrm{d}y = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)\mathrm{d}y. \end{aligned}$$

Por lo tanto, si X y Y son variables aleatorias continuas independientes entonces, la fórmula de *convolución* es la función de densidad de la suma $Z = X + Y$:

$$f(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)\mathrm{d}x = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)\mathrm{d}y; \quad z \in \mathbf{R}. \tag{4.33}$$

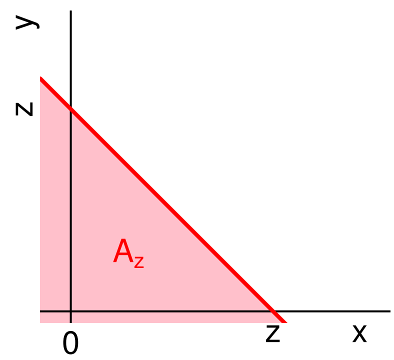


Figura 4.5. Gráfica de semiplano A_z (4.32), asociado al evento $\{X+Y \leq z\}$; para $z \in \mathbf{R}$.

Si además X y Y son variables aleatorias positivas ($P(X>0)=P(Y>0)=1$), entonces, los argumentos de los anteriores integrandos deben ser positivos: $x, y, z-x, z-y>0$. Así, la fórmula de convolución se simplifica como

$$f(z)=\int_0^z f_X(x) f_Y(z-x) \mathrm{d}x=\int_0^z f_X(z-y) f_Y(y) \mathrm{d}y ; \quad z \in \mathbf{R}.$$

Por ejemplo, si X y Y son variables aleatorias independientes, con $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, entonces

$$X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2).$$

Se verificará este resultado para el caso $\mu_1=\mu_2=0$. Al aplicar (4.33), se tiene

$$\begin{aligned} f(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) \mathrm{d}x=\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-x^2/(2\sigma_1^2)} \times \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(z-x)^2/(2\sigma_2^2)} \mathrm{d}x \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{\frac{x^2}{\sigma_1^2}+\frac{(z-x)^2}{\sigma_2^2}\right\}} \mathrm{d}x. \end{aligned}$$

Complete el trinomio cuadrado perfecto respecto de x:

$$\begin{aligned} \frac{x^2}{\sigma_1^2}+\frac{(z-x)^2}{\sigma_2^2} &= \left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) x^2-2 \frac{zx}{\sigma_2^2}+\frac{z^2}{\sigma_2^2} \\ &= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left[x^2-2 \frac{\sigma_1^2 zx}{\sigma_1^2+\sigma_2^2}+\left(\frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\right)^2\right]+\frac{z^2}{\sigma_2^2}-\frac{\sigma_1^2 z^2}{\sigma_2^2(\sigma_1^2+\sigma_2^2)} \\ &= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left[x-\frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\right]^2+\frac{z^2}{\sigma_1^2+\sigma_2^2}. \end{aligned}$$

Entonces

$$\begin{aligned}
f(z) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left(x-\frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\right)^2\right\}\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp\left(-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)\int_{-\infty}^{\infty}\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}u^2\right)\mathrm{d}u \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp\left(-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right); \quad \text{para} \quad z\in\mathbf{R}.
\end{aligned}$$

Esta es la función de densidad normal de media cero y varianza $\sigma_1^2+\sigma_2^2$. Por lo tanto

$$X+Y\sim \mathrm{N}(0,\sigma_1^2+\sigma_2^2).$$

En general, si $X_1,\ldots,X_n$ son variables aleatorias independientes, con $X_i\sim\mathrm{N}(\mu_i,\sigma_i^2)$, entonces

$$X_1+\cdots+X_n\sim\mathrm{N}(\mu_1+\cdots+\mu_n,\sigma_1^2+\cdots+\sigma_n^2).$$

Este resultado se deduce por inducción matemática. Otra vía de demostración es con las propiedades de la función generadora de momentos del Teorema 2.6.8.

Así mismo, se verifica que si $X_1,\ldots,X_n$ son variables aleatorias independientes, con $X_i\sim$ gamma(α_i,β), entonces

$$X_1+\cdots+X_n\sim\text{gamma}(\alpha_1+\cdots+\alpha_n,\beta).$$

Note que el parámetro de escala β es fijo.

De manera similar, se deduce la distribución del cociente $Z=Y/X$. Note que esta variable aleatoria está bien definida, pues $P(X=0)=0$. Al considerar (4.31), con $g(x,y)=y/x$, se tiene

$$\begin{aligned}
A_z &= \{(x,y)\in\mathbf{R}^2 : g(x,y)\le z\} = \{(x,y)\in\mathbf{R}^2 : \frac{y}{x}\le z\} \\
&= \{(x,y)\in\mathbf{R}^2 : y\ge xz,\ x<0\}\cup\{(x,y)\in\mathbf{R}^2 : y\le xz,\ x>0\}
\end{aligned}$$

y

$$\begin{aligned}
F(z) &= P(Z\le z) = P\left(\frac{Y}{X}\le z\right) \\
&= P(Y\ge zX, X<0)+P(Y\le zX, X>0) \\
&= \int_{-\infty}^{0}\int_{xz}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x + \int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)\mathrm{d}y\mathrm{d}x.
\end{aligned}$$

Por lo que

$$\begin{aligned}
f(z) = F'(z) &= \frac{\mathrm{d}}{\mathrm{d}z}\left\{\int_{-\infty}^{0}\int_{xz}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x + \int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)\mathrm{d}y\mathrm{d}x\right\} \\
&= \int_{-\infty}^{0}\frac{\mathrm{d}}{\mathrm{d}z}\int_{xz}^{\infty} f(x,y)\mathrm{d}y\mathrm{d}x + \int_{0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}z}\int_{-\infty}^{xz} f(x,y)\mathrm{d}y\mathrm{d}x \\
&= \int_{-\infty}^{0}(-x)f_{X,Y}(x,xz)\mathrm{d}x + \int_{0}^{\infty} x f_{X,Y}(x,xz)\mathrm{d}x \\
&= \int_{-\infty}^{\infty}|x| f_{X,Y}(x,xz)\mathrm{d}x.
\end{aligned}$$

Si además X y Y son variables aleatorias independientes, entonces

$$f(z) = \int_{-\infty}^{\infty}|x| f_X(x) f_Y(xz)\mathrm{d}x; \quad \text{para} \quad z \in \mathbf{R}. \tag{4.34}$$

En particular, para variables aleatorias continuas positivas, resulta

$$f(z) = \int_{0}^{\infty} x f_X(x) f_Y(xz)\mathrm{d}x; \quad \text{para} \quad z > 0.$$

Ejemplo 4.4.1. Si X y Y son variables aleatorias independientes, con $X \sim \text{gamma}(\alpha_1, \lambda)$ y $Y \sim \text{gamma}(\alpha_2, \lambda)$, entonces

$$f_{Y/X}(z) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\frac{z^{\alpha_2-1}}{(z+1)^{\alpha_1+\alpha_2}}; \quad z > 0. \tag{4.35}$$

De hecho, por (4.34), se tiene

$$\begin{aligned}
f_{Y/X}(z) &= \int_{-\infty}^{\infty}|x| f_X(x) f_Y(xz)\mathrm{d}x = \int_{0}^{\infty} x\frac{\lambda^{\alpha_1}x^{\alpha_1-1}}{\Gamma(\alpha_1)}e^{-\lambda x}\cdot\frac{\lambda^{\alpha_2}(xz)^{\alpha_2-1}}{\Gamma(\alpha_2)}e^{-\lambda(xz)}\mathrm{d}x \\
&= z^{\alpha_2-1}\int_{0}^{\infty}\frac{\lambda^{\alpha_1+\alpha_2}x^{\alpha_1+\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}e^{-\lambda(z+1)x}\mathrm{d}x \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\frac{z^{\alpha_2-1}}{(z+1)^{\alpha_1+\alpha_2}}\int_{0}^{\infty}\frac{[\lambda(z+1)]^{\alpha_1+\alpha_2}x^{\alpha_1+\alpha_2-1}}{\Gamma(\alpha_1+\alpha_2)}e^{-\lambda(z+1)x}\mathrm{d}x \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\frac{z^{\alpha_2-1}}{(z+1)^{\alpha_1+\alpha_2}}; \quad \text{para} \quad z > 0.
\end{aligned}$$

Ejemplo 4.4.2. Encuentre la función de densidad de $Z = Y^2/X^2$, donde X y Y son variables aleatorias independientes de distribución idéntica normal estándar $\mathrm{N}(0,1)$. Primero obtenga la función de densidad de $W = X^2$. Note que $g(x) = x^2$ no es una función 1-1. Sin embargo, se replicarán los pasos de la demostración del Teorema de transformación 3.6.1. Con esta idea, se tiene

$$\begin{aligned}
F(w) = P(W \leq w) &= P(X^2 \leq w) \\
&= P(|X| \leq \sqrt{w}) = P(-\sqrt{w} \leq X \leq \sqrt{w}) \\
&= \Phi(\sqrt{w}) - \Phi(-\sqrt{w}) \\
&= \Phi(\sqrt{w}) - [1 - \Phi(\sqrt{w})] \\
&= 2\Phi(\sqrt{w}) - 1; \quad w > 0.
\end{aligned}$$

Aquí se usó la identidad (3.35). Al derivar respecto de $w > 0$, se tiene

$$\begin{aligned} f(w) = F'(w) &= \frac{\mathrm{d}}{\mathrm{d}w}\left\{2\Phi(\sqrt{w}) - 1\right\} \\ &= 2\frac{1}{2\sqrt{w}}\Phi'(\sqrt{w}) = \frac{1}{\sqrt{w}}\phi(\sqrt{w}) \\ &= \frac{1}{\sqrt{w}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\sqrt{w})^2} = \frac{w^{-1/2}}{\sqrt{2\pi}}e^{-w/2}. \end{aligned}$$

Esta es la función de densidad gamma de parámetros forma $\alpha = 1/2$ y tasa $\lambda = 1/2$. Recuerde que $\Gamma(1/2) = \sqrt{\pi}$. Las variables aleatorias X^2 y Y^2 comparten función de densidad:

$$f_{X^2}(w) = f_{Y^2}(w) = \frac{w^{-1/2}}{\sqrt{2\pi}}e^{-w/2}; \quad w > 0. \tag{4.36}$$

Como se describe más adelante, esta es la distribución “ji-cuadrada con un grado de libertad”. Note que la función de densidad no está acotada superiormente en su punto moda $x = 0$. Por último, la función de densidad del cociente $Z = Y^2/X^2$ se obtiene de (4.35)

$$\begin{aligned} f(z) &= \frac{\Gamma(1/2+1/2)}{\Gamma(1/2)\Gamma(1/2)}\frac{z^{1/2-1}}{(z+1)^{1/2+1/2}} \\ &= \frac{\Gamma(1)}{(\sqrt{\pi})^2}\frac{z^{-1/2}}{z+1} \\ &= \frac{1}{\pi}\frac{1}{(z+1)\sqrt{z}}; \quad z > 0. \end{aligned}$$

Distribuciones ji-cuadrada y de Fisher. La distribuciones de probabilidad ji-cuadrada y F de Fisher se utilizan para realizar inferencia estadística de parámetros de poblaciones con distribución normal.

La distribución χ^2, con $n > 0$ *grados de libertad*, es el caso particular de la distribución gamma$(\alpha = n/2, \lambda = 1/2)$:

$$f(x) = \frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-x/2}; \quad x > 0; \quad n > 0.$$

El parámetro n es cualquier número real positivo. En este caso

$$\begin{aligned} \mathrm{E}\, X &= \frac{\alpha}{\lambda} = \frac{n/2}{1/2} = n, \\ \mathrm{V}\, X &= \frac{\alpha}{\lambda^2} = \frac{n/2}{1/2^2} = 2n. \end{aligned}$$

Por ejemplo, si $X_1, \ldots, X_n$ son n variables aleatorias independientes con distribución normal de media μ y varianza σ^2, entonces de (4.36) se tiene

$$\begin{aligned} Y_i &= \left(\frac{X_i - \mu}{\sigma}\right)^2 \\ &\sim \text{gamma}(\alpha = 1/2, \lambda = 1/2) \\ &\sim \chi^2_1. \end{aligned}$$

Pero como la distribución gamma es cerrada bajo suma de variables aleatorias independientes, entonces

$$\begin{aligned} Y_1 + \cdots + Y_n &= \left(\frac{X_1-\mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n-\mu}{\sigma}\right)^2 \\ &\sim \text{gamma}(\alpha = n/2, \lambda = 1/2) \\ &\sim \chi^2_n. \end{aligned}$$

Por otra parte, la distribución *F-Fisher*, de $m > 0$ y $n > 0$ *grados de libertad*, se define como la distribución que corresponde al cociente de dos variables aleatorias χ^2, descontadas por sus respectivos grados de libertad:

$$W \stackrel{d}{=} \frac{X/m}{Y/n} \sim F_{m,n}.$$

La forma de la densidad de esta distribución de probabilidad se obtiene de (4.35). De hecho

$$\begin{aligned} X &\sim \chi^2_m \sim \text{gamma}(\alpha = m/2, \lambda = 1/2), \\ Y &\sim \chi^2_n \sim \text{gamma}(\alpha = n/2, \lambda = 1/2), \end{aligned}$$

implica

$$f_{X/Y}(z) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{z^{m/2-1}}{(z+1)^{(m+n)/2}}; \quad z > 0.$$

Pero como

$$W = \frac{X/m}{Y/n} = \frac{n}{m}\frac{X}{Y},$$

entonces

$$\begin{aligned} f(w) &= \frac{m}{n} f_{X/Y}\left(\frac{m}{n}w\right) = \frac{m}{n}\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{\{(m/n)w\}^{m/2-1}}{((m/n)w+1)^{(m+n)/2}} \\ &= \left(\frac{m}{n}\right)^{m/2} \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{w^{m/2-1}}{(1+(m/n)w)^{(m+n)/2}}; \quad w > 0. \end{aligned}$$

4.4.1. Suma finita de variables aleatorias

4.5. Ejercicios

1. Compruebe que (X, Y) es un vector aleatorio en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, si sólo si
$$\{\omega \in \Omega : X(\omega) \le x, Y(\omega) \le y\} \in \mathcal{A}; \quad \text{para todo} \quad x, y \in \mathbf{R}.$$

2. Sea (X, Y) un vector aleatorio con función de distribución conjunta $F(x, y) = P(X \le x, Y \le y)$; para $x, y \in \mathbf{R}$. Demuestre la equivalencia de las siguientes afirmaciones

    a) (X, Y) es un vector aleatorio continuo *b*) X y Y son variables aleatorias continuas

c) $F(x, y)$ es función continua en el plano $\mathbf{R}^2$

d) $F(x-, y-) = F(x, y)$; para todo $x, y \in \mathbf{R}$

3. Sea (X, Y) un vector aleatorio continuo con función de densidad conjunta $f(x, y)$. Sobre la independencia de las variables a aleatorias X y Y, demostrar la equivalencia de las expresiones (2.46) y (4.12).

4. Sea (X, Y) un vector aleatorio de distribución normal bivariada. Verifique que la distribución condicional de la variable aleatoria X, dado $Y = y \in \mathbf{R}$, es normal, según (4.25).

5. Sean las variables aleatorias independientes X y Y, tal que X tiene distribución gamma $\text{gamma}(\alpha, \lambda = 1)$ y Y es de distribución exponencial $\exp(\lambda = 1)$. Defina $R = X/(X + Y)$ y $S = X + Y$. Véase [3].

   a) Compruebe que R y S son variables aleatorias independientes

   b) Verifique que la función de densidad de R es

   $$f(r) = \alpha r^{\alpha-1}; \quad \text{para} \quad 0 < r < 1.$$

   ¿Qué familia de distribuciones de probabilidad incluye la de esta variable aleatoria?

   c) Verifique que R^α tiene distribución uniforme continua en $(0, 1)$.

6. Práctica. Sea $X = (X_1, X_2)$ un vector aleatoria de distribución normal bivariada con vector de medias cero y matriz de covarianza Σ, determinada por los valores de los parámetros σ_1, σ_2 y ρ, de cada uno de los nueve ejemplos de la Figura 4.3.

   a) Para cada una de las probabilidades de cobertura $\alpha = 0.9, 0.95, 0.99$, obtenga la curva de nivel $c = c(\alpha) > 0$, de la función de densidad (4.20).

   b) Obtenga las gráficas correspondientes, con una unidad de medida común en los ejes coordenados.

   c) Por último, verifique los resultados obtenidos al simular $n = 5000$ vectores aleatorios (X_1, X_2). Luego, compare las probabilidades de cobertura establecidas con sus respectivas frecuencias observadas. Sugerencia: realice los siguientes pasos.

   - En el caso normal estándar con independencia, considere el círculo en el plano de centro $(0, 0)$ y radio $r > 0$:

   $$R = \{(x_1, x_2) \in \mathbf{R}^2 : x_1^2 + x_2^2 < r^2\}. \tag{4.37}$$

   Para $0 < \alpha < 1$, encuentre el radio $r = r(\alpha)$, tal que

   $$P((X_1, X_2) \in R) = P(X_1^2 + X_2^2 < r^2) = \alpha.$$

   ¿Qué distribución tiene $X_1^2 + X_2^2$?

- Para el resto de los casos, considere la *factorización de Cholesky* de la matriz Σ:

$$\Sigma = AA' = \begin{pmatrix} \sqrt{1-\rho^2}\sigma_1 & \rho\sigma_1 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \sqrt{1-\rho^2}\sigma_1 & 0 \\ \rho\sigma_1 & \sigma_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

con

$$A = \begin{pmatrix} \sqrt{1-\rho^2}\sigma_1 & \rho\sigma_1 \\ 0 & \sigma_2 \end{pmatrix}.$$

Esta factorización es tal que A sea una matriz triangular superior. En este sentido

$$Y = AX \sim \mathrm{N}(0, AI_2A') \sim \mathrm{N}(0, AA') \sim \mathrm{N}(0, \Sigma)$$

y

$$P(Y \in AR) = P(AX \in AR) = P(X \in R) = \alpha,$$

con

$$AR = \{y = Ax : x \in R\} = \left\{(y_1, y_2) \in \mathbf{R}^2 : \frac{y_1^2}{\sigma_1^2} - 2\rho\frac{y_1y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2} < (1-\rho^2)r^2\right\}.$$

En la última igualdad se usa (4.37) y

$$x = A^{-1}y = \frac{1}{\sqrt{1-\rho^2}\sigma_1\sigma_2}\begin{pmatrix} \sigma_2 & -\rho\sigma_1 \\ 0 & \sqrt{1-\rho^2}\sigma_1 \end{pmatrix} y = \frac{1}{\sqrt{1-\rho^2}\sigma_1\sigma_2}\begin{pmatrix} \sigma_2y_1 - \rho\sigma_1y_2 \\ \sqrt{1-\rho^2}\sigma_1y_2 \end{pmatrix}.$$

Capítulo 5

Convergencia de variables aleatorias

En este capítulo se describen los conceptos básicos de convergencia de variables aleatorias. En particular se muestra una versión básica del teorema del límite central.

5.1. Convergencia de variables aleatorias

En teoría de probabilidad hay varios conceptos de convergencia de variables aleatorias. En esta sección se abordarán tres de ellos: convergencia en distribución o débil, convergencia en probabilidad y convergencia casi segura o fuerte.

La convergencia casi segura es análogo al de límite de una sucesión de funciones reales, pues una variable aleatoria es formalmente una función real; véase [12, Definición 7.1]. La convergencia en probabilidad es un concepto más débil, en el sentido de que se deduce ante una convergencia casi segura. A su vez, la convergencia en probabilidad implica la de en distribución. Por lo cual se le dice convergencia débil.

Sea la colección de variables aleatorias X y $\{X_n\}_{n\geq 1}$, definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se dice que la sucesión $\{X_n\}$ converge *casi seguramente* a X, si $A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ es un evento seguro:

$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1. \tag{5.1}$$

Se denota como cualquiera de las siguientes expresiones

$$\begin{aligned}
P(\lim_{n\to\infty} X_n = X) &= 1,\\
\lim_{n\to\infty} X_n &= X; \quad \text{c.s.},\\
X_n &\to X; \quad n\to\infty; \quad \text{c.s.},\\
X_n &\overset{\text{c.s.}}{\to} X; \quad n\to\infty,\\
X_n &\overset{\text{c.s.}}{\to} X.
\end{aligned}$$

Si $\omega \in A^c$, entonces la sucesión de números reales $\{X_n(\omega)\}_{n\geq 1}$ no tiene limite o no converge a $X(\omega)$. Sin embargo, el evento A^c no ocurrirá en la práctica, pues tiene probabilidad cero. Por otro lado, note que

$$X_n \overset{\text{c.s.}}{\to} X, \quad \text{si y sólo si} \quad X_n - X \overset{\text{c.s.}}{\to} 0; \quad \text{cuando} \quad n\to\infty.$$

La colección de variables aleatorias $\{X_n\}_{n\geq 1}$, converge *en probabilidad* a la variable aleatoria X si, para cualquier $\varepsilon > 0$, se satisface

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \tag{5.2}$$

Por el Ejercicio 5.3.2, esta condición se reemplaza con cualquiera de las siguientes afirmaciones equivalentes:

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0, \tag{5.3}$$

$$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1, \tag{5.4}$$

$$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1. \tag{5.5}$$

La convergencia en probabilidad se denota con cualquiera de las siguientes expresiones

$$\begin{aligned} \lim_{n\to\infty} X_n = X; &\quad \text{en probabilidad,} \\ X_n \to X; &\quad n\to\infty; \quad \text{en probabilidad,} \\ X_n \overset{\mathrm{p}}{\to} X; &\quad n\to\infty, \\ X_n \overset{\mathrm{p}}{\to} X&. \end{aligned}$$

Note que

$$X_n \overset{\mathrm{p}}{\to} X, \quad \text{si y sólo si} \quad X_n - X \overset{\mathrm{p}}{\to} 0; \quad \text{cuando} \quad n\to\infty.$$

La convergencia casi segura implica la de en probabilidad, como se afirma en la Proposición 5.1.1, más adelante. Si bien el reciproco no es cierto, si $X_n \overset{\mathrm{p}}{\to} X$, entonces existe una subsucesión $\{X_{k_n}\}_{n\geq 1}$, tal que $X_{k_n} \overset{\mathrm{c.s.}}{\to} X$; $n\to\infty$.

Sean $F(x)$ y $F_n(x)$; $n\geq 1$, las funciones de distribución de las variables aleatorias X y X_n; $n\geq 1$, respectivamente. Entonces $\{X_n\}_{n\geq 1}$ converge *en distribución* o *débilmente* a la variable aleatoria X si

$$\lim_{n\to\infty} F_n(x) = F(x); \tag{5.6}$$

para todo punto de continuidad $x \in \mathbf{R}$ de $F(\cdot)$. Por el Ejercicio 5.3.3, esta condición se reemplaza con cualquiera de las siguientes afirmaciones equivalentes::

$$\lim_{n\to\infty} P(X_n < x) = P(X < x), \tag{5.7}$$

$$\lim_{n\to\infty} P(X_n \geq x) = P(X \geq x), \tag{5.8}$$

$$\lim_{n\to\infty} P(X_n > x) = P(X > x). \tag{5.9}$$

La convergencia en distribución se denota con cualquiera de las siguientes expresiones

$$\begin{aligned} \lim_{n\to\infty} X_n = X; &\quad \text{en distribución,} \\ X_n \to X; &\quad n\to\infty; \quad \text{en distribución,} \\ X_n \overset{\mathrm{d}}{\to} X; &\quad n\to\infty, \\ X_n \overset{\mathrm{d}}{\to} X&, \\ F_n \overset{\mathrm{d}}{\to} F&. \end{aligned}$$

En contraste con las convergencias casi segura y en probabilidad, aquí no hay una afirmación equivalente de la distribución límite de $X_n - X$. La convergencia en probabilidad implica la de en distribución, como se afirma en la Proposición 5.1.1, más adelante. Si bien el reciproco no es cierto, si $X_n \xrightarrow{\text{d}} X$, entonces existe una subsucesión $\{X_{k_n}\}_{n\geq 1}$, tal que $X_{k_n} \xrightarrow{\text{p}} X$; $n \to \infty$.

Proposición 5.1.1.
1. Si $X_n \xrightarrow{\text{c.s.}} X$, entonces $X_n \xrightarrow{\text{p}} X$; cuando $n \to \infty$. El recíproco no es cierto.
2. Si $X_n \xrightarrow{\text{p}} X$, entonces $X_n \xrightarrow{\text{d}} X$; cuando $n \to \infty$. El recíproco no es cierto.
3. $X_n \xrightarrow{\text{p}} c$, si y sólo si $X_n \xrightarrow{\text{d}} c$; cuando $n \to \infty$, donde c es una constante.

Demostración. Sean los eventos

$$\begin{aligned} A &= \{\lim_{n\to\infty} X_n = X\} = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}, \\ B &= \bigcup_{m\geq 1}\bigcap_{n\geq m}\{|X_n - X| < \varepsilon\} = \bigcup_{m\geq 1}\bigcap_{n\geq m}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}, \end{aligned}$$

donde $\varepsilon > 0$ es fijo. Se verificará la igualdad de ambos eventos. Si $\omega \in A$, entonces existe $m \geq 1$ tal que $|X_n(\omega) - X(\omega)| < \varepsilon$; para todo $n \geq m$. Así que

$$\omega \in \bigcap_{n\geq m}\{|X_n - X| < \varepsilon\} \subset B.$$

Recíprocamente, si $\omega \in B$, entonces

$$\omega \in \bigcap_{n\geq m}\{|X_n - X| < \varepsilon\}; \quad \text{para cierto} \quad m \geq 1,$$

$$|X_n(\omega) - X(\omega)| < \varepsilon; \quad \text{para todo} \quad n \geq m \quad \text{y cierto} \quad m \geq 1.$$

Como $\varepsilon > 0$ es arbitrario, se obtiene $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ y $\omega \in A$.
1. Ahora suponga que $X_n \xrightarrow{\text{c.s.}} X$; $n \to \infty$. Como el evento B es una unión creciente de eventos y de la propiedad de continuidad de la probabilidad (1.23), se tiene

$$1 = P(A) = P(B) = \lim_{m\to\infty} P\left(\bigcap_{n\geq m}\{|X_n - X| < \varepsilon\}\right) \leq \lim_{m\to\infty} P(\{|X_m - X| < \varepsilon\}).$$

Se concluye $X_n \xrightarrow{\text{p}} X$; $n \to \infty$.
2. Considere $X_n \xrightarrow{\text{p}} X$; $n \to \infty$. Sin pérdida de generalidad, asuma que $X = 0$, con probabilidad uno. Su función de distribución es

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases}$$

Esta función sólo es continua para $x \neq 0$. Si $x > 0$, se tiene

$$P(X_n > x) \leq P(|X_n| > x) \to 0 = P(X > x); \quad n \to \infty.$$

Entonces

$$F_n(x) = P(X_n \le x) = 1 - P(X_n > x) \to 1; \quad n \to \infty.$$

De manera similar, cuando $x < 0$, se tiene $-x > 0$ y

$$F_n(x) = P(X_n \le x) = P(-X_n \ge -x) \le P(|-X_n| \ge -x) = P(|X_n| \ge -x) \to 0; \quad n \to \infty.$$

Se concluye

$$\lim_{n\to\infty} F_n(x) = F(x); \quad \text{para todo} \quad x \neq 0.$$

3. Este resultado se verifica por el Ejercicio 5.3.4. □

Un contra ejemplo de la Proposición 5.1.1.2 se obtiene al considerar las sucesión de variables aleatorias independientes e idénticamente distribuidas $\{X_k\}_{k\ge 0}$, de distribución binomial de parámetros $n = 5$ y $p = 18/38$. Entonces

$$X_0 \stackrel{d}{=} X_1 \cdots \stackrel{d}{=} X_k \stackrel{d}{=} \cdots$$

Así, $X_k \xrightarrow{\mathrm{d}} X_0$, aunque no hay una convergencia en probabilidad ni casi segura.

Por el Teorema 2.7.1, la aproximación de Poisson es otro ejemplo de convergencia en distribución. Si $X_n \sim \mathrm{bin}(n, p_n)$; $n \ge 1$ y $\mu = \lim_{n\to\infty} np_n > 0$, entonces $X_n \xrightarrow{\mathrm{d}} X$, donde X tiene distribución de Poisson de media μ:

$$\lim_{n\to\infty} P(X_n = x) = P(X = x) = e^{-\mu}\frac{\mu^x}{x!}; \quad \text{para} \quad x = 0, 1, \ldots \tag{5.10}$$

Por lo tanto

$$\begin{aligned}\lim_{n\to\infty} P(X_n \le x) &= \lim_{n\to\infty} \sum_{y=0}^{x} P(X_n = y) = \sum_{y=0}^{x} \lim_{n\to\infty} P(X_n = y) \\ &= \sum_{y=0}^{x} P(X = y) = P(X \le x); \quad x = 0, 1, \ldots \end{aligned} \tag{5.11}$$

Este resultado aplica también para cualquier punto de la recta real. En general, como en (5.10)-(5.11), la convergencia en distribución de variables aleatorias discretas equivale al límite de las respectivas funciones de densidad; véase Ejercicio 5.3.5.

Ejemplo 5.1.2. Sea $X_1, X_2, \ldots$, una sucesión de variables aleatorias independientes de distribución idéntica $\mathrm{N}(\mu, \sigma^2)$. Entonces

$$\sum_{i=1}^{n} X_i \sim \mathrm{N}(n\mu, n\sigma^2),$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Note que la varianza de la media muestral converge a cero, cuando $n \to \infty$. Por el ejercicio 5.3.6, esto implica

$$\bar{X}_n \xrightarrow{\mathrm{d}} \mu; \quad n \to \infty.$$

De hecho, si $F_n(x)$ es la función de distribución de la media muestral $\bar{X}_n$, entonces

$$F_n(x) = P(\bar{X}_n \leq x) = P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{x-\mu}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right); \quad x \in \mathbf{R}.$$

Así

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} \Phi\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right) = \lim_{n\to\infty} \Phi(\sqrt{n}\,\mathrm{signo}(x-\mu)) = \begin{cases} 1 & x > \mu \\ 1/2 & x = \mu \\ 0 & x < \mu. \end{cases}$$

Por otra parte, la función de distribución de la variable aleatoria constante $X_0 \equiv \mu$ es

$$F_0(x) = \begin{cases} 1 & x \geq \mu \\ 0 & x < \mu. \end{cases}$$

Por lo que

$$\lim_{n\to\infty} F_n(x) = F_0(x); \quad x \neq \mu$$

y

$$\bar{X}_n \xrightarrow{\mathrm{d}} \mu; \quad n \to \infty.$$

En la Figura 5.1 se aprecian las gráficas de la función de densidad y de distribución de $\bar{X}_n$, para diferentes valores de n, con media $\mu = 13.75$ y desviación estándar $\sigma = 2$.

5.2. Teorema del límite central

Sea $X_1, X_2, \ldots, X_n$ una sucesión de variables aleatorias independientes idénticamente distribuidas, de una población con media $\mu \in \mathbf{R}$ y varianza finita $\sigma^2 > 0$. Recuerde que la media muestral

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

satisface

$$\mathrm{E}\,\bar{X}_n = \mu, \quad \mathrm{V}\,\bar{X}_n = \frac{\sigma^2}{n} \quad \text{y} \quad \mathrm{d.e.}(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}.$$

Su versión estándar es

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Teorema 5.2.1 (teorema del límite central)**.** La sucesión de variables aleatorias $\{Z_n\}_{n\geq 1}$ converge en distribución a una variable aleatoria Z de distribución normal estándar:

$$\lim_{n\to\infty} P(Z_n \leq z) = P(Z \leq z) = \Phi(z); \quad \text{para todo} \quad z \in \mathbf{R}.$$

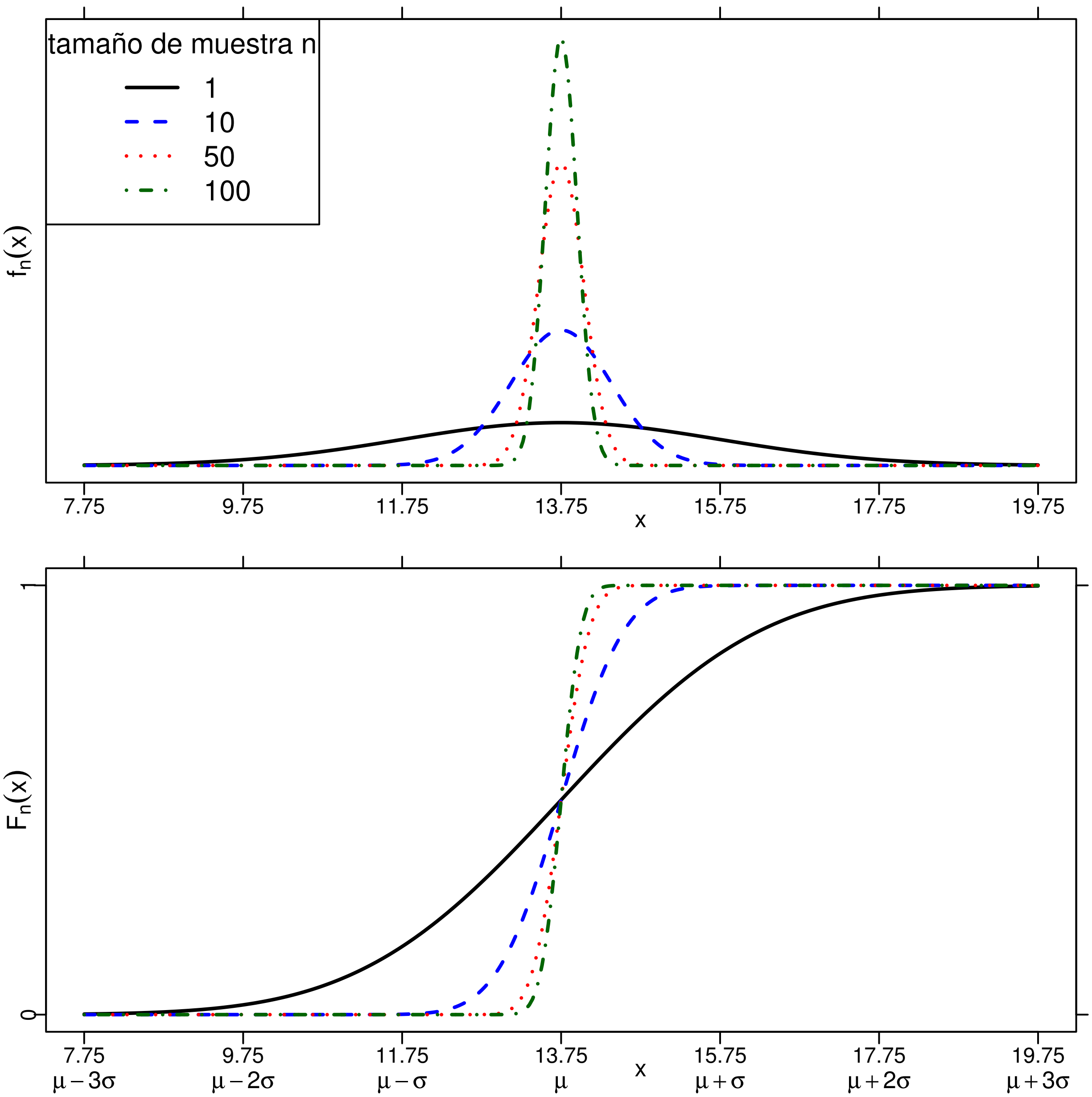


Figura 5.1. [arriba] Gráficas de funciones de densidad de media muestral $\bar{X}_n$, de una población $N(\mu = 13.75, \ \sigma^2 = 4)$; para tamaños de muestra $n = 1, 10, 50, 100$. [abajo] Respectivas gráficas de funciones de distribución.

Demostración. Se probará sólo el caso donde esté bien definida la función generadora de momentos de la variable aleatoria X_1: $M(t) = \operatorname{E} e^{tX_1} < \infty$; para $t \in (-h, h)$ y cierto $h > 0$. Sin pérdida de generalidad, asuma $\mu = 0$ y $\sigma^2 = 1$. Entonces

$$Z_n = \frac{\bar{X}_n}{1/\sqrt{n}} = \sqrt{n}\frac{\sum_{i=1}^n X_i}{n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i.$$

La función generadora de momentos de Z_n es

$$\begin{aligned} M_n(t) &= \operatorname{E} e^{tZ_n} = \operatorname{E} e^{t(X_1+\cdots+X_n)/\sqrt{n}} \\ &= \operatorname{E}[e^{(t/\sqrt{n})X_1}\cdots e^{(t/\sqrt{n})X_n}] = \operatorname{E} e^{(t/\sqrt{n})X_1}\cdots \operatorname{E} e^{(t/\sqrt{n})X_n} \\ &= [\operatorname{E} e^{(t/\sqrt{n})X_1}]^n = [M(t/\sqrt{n})]^n \quad \text{para} \quad t \in \mathbf{R}. \end{aligned}$$

Por la expansión en serie de Taylor de $M(t)$, se tiene

$$\begin{aligned} M\left(\frac{t}{\sqrt{n}}\right) &= M(0) + M'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}M''(0)\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right) \\ &= 1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) = 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right). \end{aligned}$$

Aquí se usó

$$M(0) = 1, \quad M'(0) = \operatorname{E} X = 0 \quad \text{y} \quad M''(0) = \operatorname{E} X^2 = \operatorname{V} X = 1.$$

Al considerar (21), para $t \in \mathbf{R}$, se tiene

$$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n = e^{t^2/2} = M_Z(t).$$

□

Una consecuencia del teorema del límite central es la *ley débil de los grandes números*:

$$\bar{X}_n \xrightarrow{\mathrm{d}} \mu,$$

donde $\mu = \operatorname{E} X$, es la media poblacional; véase el Ejercicio 5.3.6. Se verificará la convergencia en probabilidad. Si $\varepsilon > 0$, entonces

$$\begin{aligned} \lim_{n\to\infty} P(|\bar{X}_n - \mu| \le \varepsilon) &= \lim_{n\to\infty} P\left(\frac{|\bar{X}_n - \mu|}{\sigma/\sqrt{n}} \le \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = \lim_{n\to\infty} P\left(|Z_n| \le \frac{\sqrt{n}\varepsilon}{\sigma}\right) \\ &= \lim_{n\to\infty} P\left(-\frac{\sqrt{n}\varepsilon}{\sigma} \le Z_n \le \frac{\sqrt{n}\varepsilon}{\sigma}\right) = \lim_{n\to\infty}\left[\Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\sqrt{n}\varepsilon}{\sigma}\right)\right] \\ &= \Phi(\infty) - \Phi(-\infty) = 1. \end{aligned}$$

Este resultado generaliza al obtenido en el Ejemplo 5.1.2, para el caso de una población de distribución normal.

5.3. Ejercicios

1. Práctica. Lance $n = 10$ dados y registre la suma Y de los números de las caras superiores: $Y = X_1 + \cdots + X_{10}$, donde $X_1, \ldots, X_{10}$ es una muestra aleatoria de tamaño $n = 10$, de variables aleatorias independientes de distribución uniforme discreta en $\{1, 2, 3, 4, 5, 6\}$.

   a) Identifique el conjunto de valores posibles de la variable aleatoria Y

   b) Obtenga la esperanza, varianza y desviación estándar de Y

   c) Repita el experimento anterior $m = 80$ veces. Con esta muestra aleatoria de tamaño $m = 80$, obtenga un histograma de frecuencias absolutas. Considere intervalos de clase de longitud uno o dos.

   d) Realice el anterior histograma con ajuste de densidad normal. Compruebe que la forma de campana de Gauss de la función de densidad normal explica la tendencia de las frecuencias absolutas

   e) Obtenga un intervalo de predicción aproximado para Y de cobertura 90 %:

   $$Y \in \mathrm{E}\, Y \pm z_{\alpha/2}\, \mathrm{d.\,e.}(Y),$$

   donde z_α denota el α-cuantil superior de la distribución normal estándar; para $0 < \alpha < 1$. En este caso, $\alpha = 1 - 0.9 = 0.1$ y $\alpha/2 = 0.05$.

   f) Ahora obtenga un intervalo de predicción aproximado para Y de cobertura 95 %.

2. Demuestra la equivalencia de las expresiones (5.2)-(5.5), para obtener la convergencia en probabilidad.

3. Demuestra la equivalencia de las expresiones (5.6)-(5.9), para obtener la convergencia en distribución.

4. Demuestre la equivalencia entre las convergencias en probabilidad y en distribución, cuando la distribución límite corresponde a la de una constante real c:

   $$X_n \xrightarrow{\mathrm{p}} c, \quad \text{si y sólo si} \quad X_n \xrightarrow{\mathrm{d}} c.$$

5. Sea $\{X_n\}_{n \geq 0}$ una sucesión de variables aleatorias discretas, con conjunto de valores posibles en los enteros no negativos. Denote como $f_n(x)$ y $F_n(x)$; $n \geq 0$, sus respectivas funciones de densidad y de distribución. Verifique la equivalencia de las siguientes afirmaciones.

   a) $X_n \xrightarrow{\mathrm{d}} X_0$

   b) $f_n(x) \to f_0(x)$; para todo $x = 0, 1, \ldots$

   c) $F_n(x) \to F_0(x)$; para todo $x \in \mathbf{R}$. Nota: este límite aplica incluso en los puntos de discontinuidad de $F_0(x)$.

6. Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias tal que

   $$\lim_{n \to \infty} \mathrm{E}\, X_n = \mu \in \mathbf{R} \quad \text{y} \quad \lim_{n \to \infty} \mathrm{V}\, X_n = 0.$$

   Pruebe que $X_n \xrightarrow{\mathrm{d}} \mu$. Sugerencia: use la desigualdad de Chebyshev.

7. Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias tal que

$$X_n = \begin{cases} 0 & \text{con probabilidad } 1-1/n \\ n & \text{con probabilidad } 1/n \end{cases}.$$

Pruebe lo siguiente

a)
$$\mathrm{E}\, X_n = 1, \quad \lim_{n\to\infty} \mathrm{V}\, X_n = \infty, \quad X_n \xrightarrow{\mathrm{d}} 0 \quad \text{y} \quad X_n \xrightarrow{\mathrm{p}} 0.$$

b) X_n no converge casi seguramente a cero. Sugerencia: verifique que el evento B del inicio de la demostración de la Proposición 5.1.1 tiene probabilidad menor que uno.

8. Teorema de Slutsky. Sean $\{X_n\}_{n\geq 1}$ y $\{Y_n\}_{n\geq 1}$ dos sucesiones de variables aleatorias, X una variable aleatoria y c una constante, tales que $X_n \xrightarrow{\mathrm{d}} X$ y $Y_n \xrightarrow{\mathrm{d}} c$. Demuestre

$$\begin{aligned} X_n + Y_n &\xrightarrow{\mathrm{d}} X + c, \\ X_n Y_n &\xrightarrow{\mathrm{d}} cX, \\ \frac{X_n}{Y_n} &\xrightarrow{\mathrm{d}} \frac{X}{c}; \quad \text{siempre que} \quad c \neq 0. \end{aligned}$$

9. Demuestre la versión del Teorema de Slutsky bajo convergencia en probabilidad.

10. ¿Aplica una versión del Teorema de Slutsky bajo convergencia casi segura?

Apéndice

A Funciones reales

Sea $f : A \to B \subset \mathbf{R}$ una *función real*, la cual es de variable real o vectorial, según si $A \subset \mathbf{R}$ ó $A \subset \mathbf{R}^n$, con $n \geq 2$, respectivamente. Salvo que se mencione lo contrario, en este Apéndice, las funciones reales abordadas son de variable real.

Se dice que la función $f : A \to B$, es *inyectiva*, *invertible* o *uno a uno*, si la igualdad $f(x) = f(y)$, implica $x = y$; con $x, y \in A$. Por otro lado, la función $f(x)$ es *sobreyectiva* si su *imagen* es B:

$$f(A) = \{f(x) : x \in A\} = B.$$

La función $f(x)$ es *biyectiva* si es inyectiva y sobreyectiva. Una función real $f(x)$ es *creciente* o *creciente en el sentido amplio* si

$$f(x) \leq f(y), \quad \text{siempre que} \quad x < y.$$

Si la desigualdad de esta expresión se cumple en el sentido estricto, entonces $f(x)$ es *estrictamente creciente*. En cambio, la función $f(x)$ es *decreciente* o *decreciente en el sentido amplio* si

$$f(x) \geq f(y), \quad \text{siempre que} \quad x < y.$$

Si la desigualdad de esta expresión se cumple en el sentido estricto, entonces $f(x)$ es *estrictamente decreciente*.

Toda función real de variable real estrictamente creciente o decreciente es invertible. Por ejemplo, la función *identidad* $f(x) = x$; $x \in \mathbf{R}$, es real, de variable real, estrictamente creciente e invertible, de hecho biyectiva. Recíprocamente, si $f(x)$ es una función real definida en un intervalo, continua e invertible, entonces es estrictamente creciente o decreciente.

La función *parte entera* se define como

$$f(x) = \lfloor x \rfloor = \text{máximo entero menor o igual que } x; \quad \text{para} \quad x \in \mathbf{R}. \tag{12}$$

Por ejemplo

$$\lfloor -4 \rfloor = \lfloor -3.1 \rfloor = -4, \quad \lfloor -2.8 \rfloor = \lfloor -2.1 \rfloor = -3, \quad \lfloor \pi \rfloor = \lfloor 3.5 \rfloor = 3, \quad \lfloor 8 \rfloor = 8.$$

La función parte entera es creciente en el sentido amplio mas no invertible. Esta función es además discontinua, escalonada, continua por la derecha y tiene límite por la izquierda. Sus puntos de discontinuidad son el conjunto de los enteros $\mathbf{Z}$; véase su gráfica en la Figura 2.

Una función real $f(x)$ es *par* o *simétrica respecto de cero* si

$$f(-x) = f(x); \quad \text{para} \quad x \in \mathbf{R}.$$

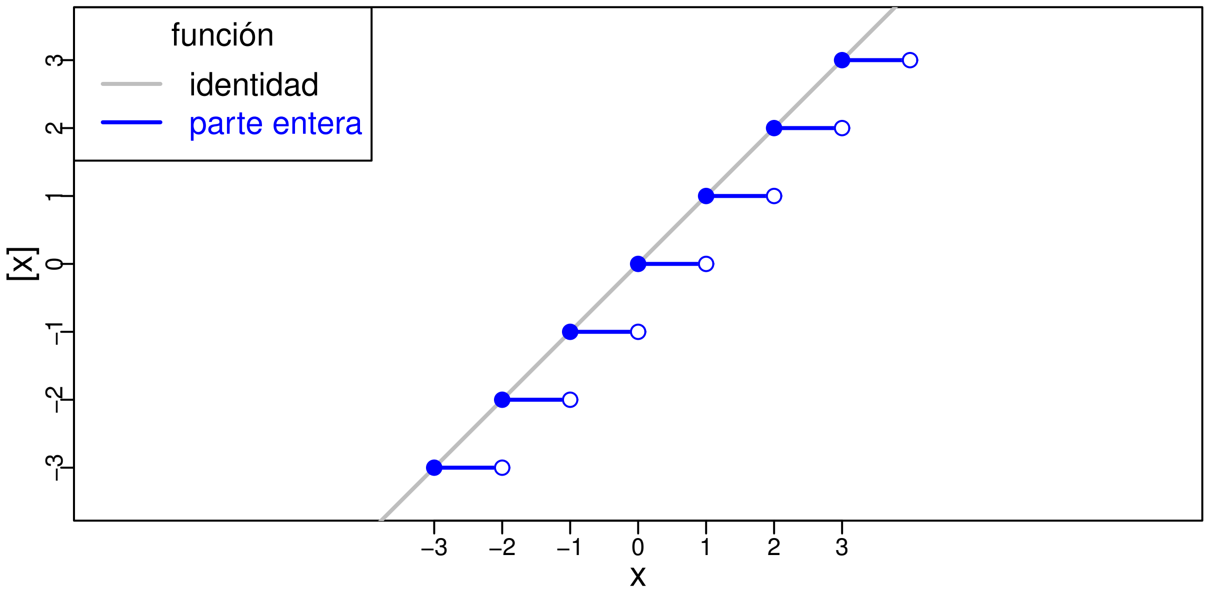


Figura 2. Gráfica de las funciones parte entera (escalonada) e identidad (continua gris).

En cambio, si

$$f(-x) = -f(x); \quad \text{para} \quad x \in \mathbf{R},$$

se dice que $f(x)$ es *impar* o *anti simétrica*. Las funciones *coseno* y *seno* son ejemplos de funciones par e impar; respectivamente:

$$\cos(-x) = \cos x \quad \text{y} \quad \operatorname{sen}(-x) = -\operatorname{sen} x; \quad x \in \mathbf{R}.$$

La función polinomial de grado impar $f(x) = x^{2k+1}$; $x \in \mathbf{R}$, con entero $k \geq 0$, es impar. En cambio, es función par de grado par $f(x) = x^{2k}$. El producto de dos funciones par ó de dos impar resulta también par. En contraste, el producto de una función par con otra impar resulta una función impar.

Sea $f(x)$ una función *integrable* en el intervalo $[-a, a]$, con $a > 0$. Si $f(x)$ es función par, entonces

$$\int_{-a}^{a} f(t)dt = 2\int_{0}^{a} f(t)dt. \tag{13}$$

En cambio

$$\int_{-a}^{a} f(t)dt = 0, \tag{14}$$

cuando $f(x)$ sea función impar. Se verificará sólo el resultado el caso impar. Como $f(-x) = -f(x)$; para todo $x \in [-a, a]$, en particular se tiene

$$f(x) = -f(-x); \quad \text{para} \quad x \in [-a, 0].$$

Así

$$\int_{-a}^{0} f(t)dt = \int_{-a}^{0} [-f(-t)]dt = -\int_{-a}^{0} f(-t)dt = \int_{a}^{0} f(u)du = -\int_{0}^{a} f(u)du.$$

Aquí se usó el cambio de variable $u = -t$. Los resultados (13) y (14) aplican también para integrales impropias, en donde $a = \infty$. Así, si $f(x)$ es función par e integrable en $\mathbf{R}$, entonces

$$\int_{-\infty}^{\infty} f(t)dt = 2\int_{0}^{\infty} f(t)dt.$$

En el caso de que la función $f(x)$ sea impar, resulta

$$\int_{-\infty}^{\infty} f(t)dt = 0. \tag{15}$$

Valor absoluto. La función *valor absoluto* se define como

$$f(x) = |x| = \begin{cases} -x & x \leq 0 \\ x & x \geq 0. \end{cases}$$

Esta función es simétrica respecto de cero y derivable para $x \neq 0$, con

$$f'(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases}$$

Las gráficas de la función valor absoluto y su derivada se muestran en la Figura 3. Para $a \geq 0$, se tienen la siguientes equivalencias

$$\begin{aligned} |x| > a \quad &\Leftrightarrow \quad x < -a \quad \text{ó} \quad x > a, \\ |x| \geq a \quad &\Leftrightarrow \quad x \leq -a \quad \text{ó} \quad x \geq a, \\ |x| < a \quad &\Leftrightarrow \quad -a < x < a, \\ |x| \leq a \quad &\Leftrightarrow \quad -a \leq x \leq a. \end{aligned}$$

Si $a < 0$, ¿cuál es la solución de cada una de las siguientes desigualdades?

$$|x| > a, \quad |x| \geq a, \quad |x| < a, \quad |x| \leq a.$$

Función gamma. La *función gamma* se define como

$$\Gamma(\alpha) = \int_{0}^{\infty} x^{\alpha-1} e^{-x} \mathrm{d}x; \quad \alpha > 0. \tag{16}$$

Esta función satisface las siguientes propiedades

$$\begin{aligned} \Gamma(1) &= 1, \\ \Gamma(2) &= 1, \\ \Gamma(n) &= (n-1)!; \quad n = 1, 2, \ldots, \\ \Gamma(1/2) &= \sqrt{\pi}. \end{aligned} \tag{17}$$

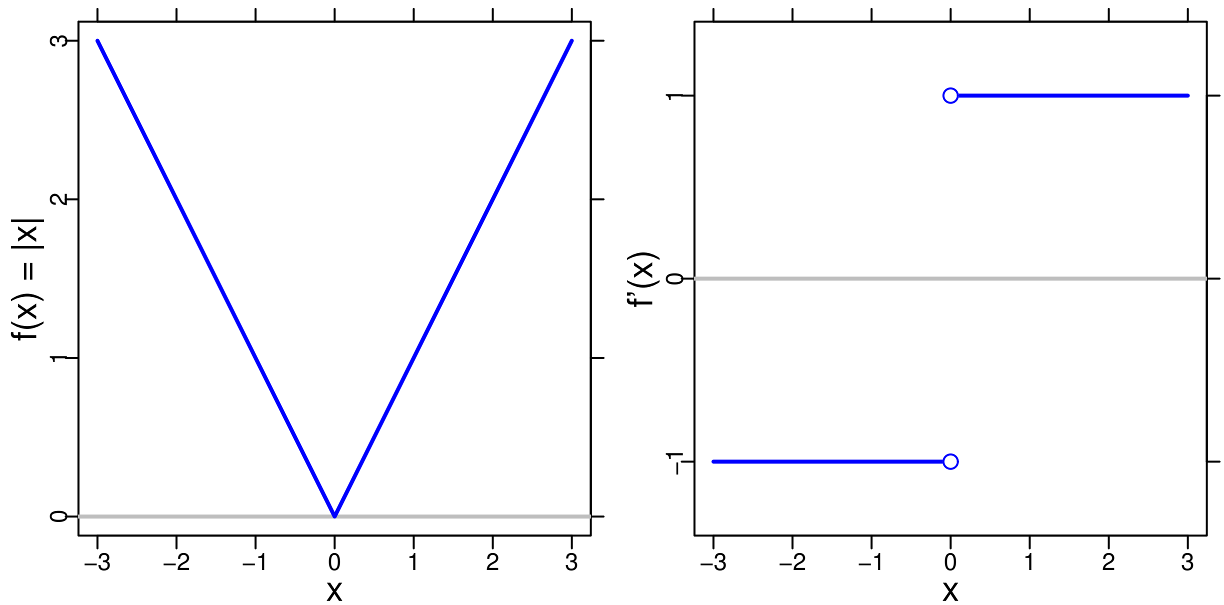


Figura 3. [izquierda] Gráfica de la función valor absoluto $f(x) = |x|$; $x \in \mathbf{R}$. [derecha] Gráfica de su función derivada.

Su fórmula recurrente es

$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha); \quad \alpha > 0. \tag{18}$$

Por ejemplo

$$\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} \mathrm{d}x = \int_0^\infty e^{-x} \mathrm{d}x = 1,$$
$$\Gamma(2) = \int_0^\infty x^{2-1} e^{-x} \mathrm{d}x = \int_0^\infty x e^{-x} \mathrm{d}x = 1.$$

La última igualdad se obtiene al recordar que tal integral es la esperanza de una variable aleatoria de distribución exponencial de parámetro tasa $\lambda = 1$. También, por la fórmula recurrente (18), se obtiene el mismo resultado

$$\Gamma(2) = \Gamma(1+1) = 1 \cdot \Gamma(1) = 1.$$

La fórmula recurrente (18) se deduce al aplicar integración por partes, $u = x^\alpha$ y $\mathrm{d}v = e^{-x}\mathrm{d}x$:

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^\infty x^{(\alpha+1)-1} e^{-x} \mathrm{d}x = \int_0^\infty x^\alpha e^{-x} \mathrm{d}x \\ &= [x^\alpha(-e^{-x})]_0^\infty - \int_0^\infty (-e^{-x})\alpha x^{\alpha-1} \mathrm{d}x \\ &= -\lim_{x\to\infty} [x^\alpha e^{-x}] + \alpha \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x \\ &= \alpha\,\Gamma(\alpha).\end{aligned}$$

Para verificar (17) se usa (3.27). Con el cambio de variable $z = \sqrt{2x}$, resulta $\mathrm{d}z = \mathrm{d}x/\sqrt{2x}$ y

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty x^{1/2-1}e^{-x}dx = \int_0^\infty e^{-x}\frac{dx}{\sqrt{x}} \\ &= \int_0^\infty e^{-z^2/2}(\sqrt{2}\mathrm{d}z) = \sqrt{\pi}\int_0^\infty \frac{2}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z \\ &= \sqrt{\pi}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z = \sqrt{\pi}.\end{aligned}$$

B Leyes de los exponentes

Las leyes de los exponentes son las siguientes. Para $a > 0$ y $x, y \in \mathbf{R}$, se tiene

$$\begin{aligned}a^x a^y &= a^{x+y}, \\ (a^x)^y &= a^{xy}, \\ a^{-x} &= \frac{1}{a^x} = \left(\frac{1}{a}\right)^x, \\ (ab)^x &= a^x b^x, \\ \left(\frac{a}{b}\right)^x &= \frac{a^x}{b^x} = a^x b^{-x}.\end{aligned}$$

Considere la función *potencia*

$$f(x) = a^x; \quad x \in \mathbf{R}, \quad \text{con} \quad a > 0. \tag{19}$$

Esta función es estrictamente creciente [decreciente] si $a > 1$ [$0 < a < 1$]. Por ejemplo, $f(x) = 2^x$; para $x \in \mathbf{R}$, es una función creciente. En cambio, es decreciente la función

$$f(x) = 2^{-x} = \frac{1}{2^x} = \left(\frac{1}{2}\right)^x; \quad x \in \mathbf{R}.$$

Otro ejemplo es la función *exponencial*

$$f(x) = \exp(x) = e^x; \quad x \in \mathbf{R}.$$

La Figura 4 muestra gráficas de la función potencia, para valores de la *base* $a = 1/3, 1/e, 2, e$. Aquí $e \approx 2.7183$, es el número de Euler.

En cambio, cuando la base no es positiva, hay restricciones en la potencia. Por ejemplo, si $a = 0$, entonces está bien definido 0^x; para $x \geq 0$:

$$0^x = \begin{cases} 1 & x = 0 \\ 0 & x > 0. \end{cases}$$

En particular, $0^0 = 1$. Cuando la base es negativa, aún hay mas restricciones para la potencia, que debe ser entera. Por ejemplo

$$\begin{array}{ccc} (-3)^{-2} = 0.1111, & (-3)^{-1} = -0.3333, & (-3)^0 = 1, \\ (-3)^1 = -3, & (-3)^2 = 9, & (-3)^3 = 27. \end{array}$$

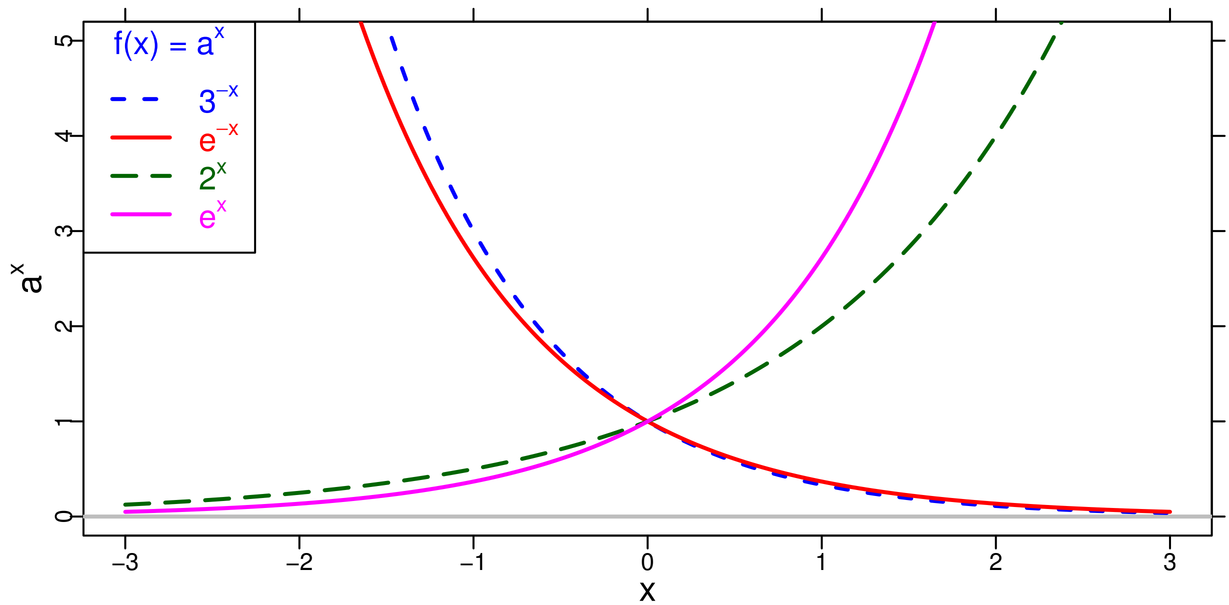


Figura 4. Gráficas de funciones potencia $f(x) = a^x$; $x \in \mathbf{R}$, con base $a = 1/3$ (discontinua azul), $1/e$ (continua roja), 2 (discontinua verde) y e (continua magenta).

Sin embargo, no son números reales sus potencias fraccionarias o irracionales:

$$(-3)^{2/5}, (-3)^{1/2}, (-3)^{\pi} \notin \mathbf{R}.$$

La función inversa de (19) es el *logaritmo base* $a > 0$:

$$\log_a(y) = x \in \mathbf{R}, \quad \text{si y sólo si} \quad a^x = y > 0.$$

Esta es una función estrictamente cóncava. Además, la función $\log_a(x)$ es estrictamente creciente [decreciente] si y sólo si lo es su inversa a^x, si y sólo si $a > 1$ $[0 < a < 1]$. El logaritmo *natural* es el de base e. Se denota como $\log x$, $\log_e x$ o $\ln x$. El logaritmo *decimal* es el de base 10, que se denota como $\log_{10} x$ o $\log x$. Cabe tener cuidado con $\log x$, pues puede ser logaritmo base e ó 10. La Figura 5 muestra gráficas de la función logaritmo, para valores de la base $a = 1/3, 1/e, 2, e$.

C Fórmula del binomio o binomio de Newton

Sean $a, b \in \mathbf{R}$ y $n \geq 0$ un entero. Entonces

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k. \tag{20}$$

Recuerde

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}; \quad \text{para} \quad 0 \leq k \leq n.$$

En particular

$$2^n = (1+1)^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-2} + \binom{n}{n-1} + \binom{n}{n}.$$

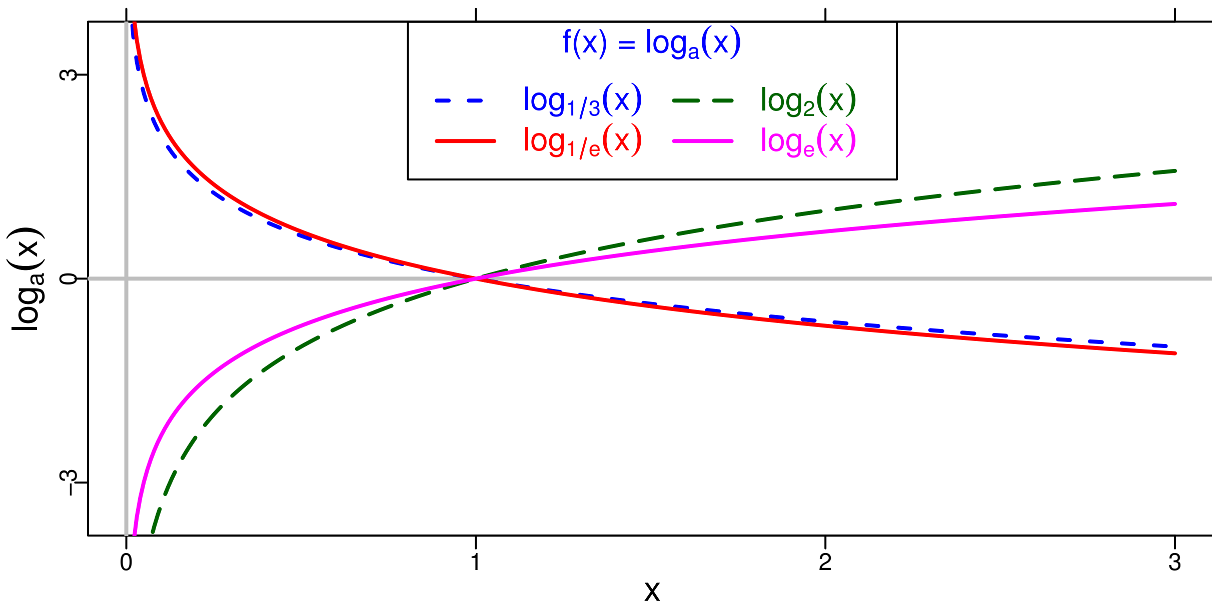


Figura 5. Gráficas de funciones logaritmo $f(x) = \log_a x$; $x > 0$, con base $a = 1/3$ (discontinua azul), $1/e$ (continua roja), 2 (discontinua verde) y e (continua magenta).

D Sucesiones y series

El número e y otros límites. El número e se define como el límite

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n .$$

Se verifica también

$$e^{ab} = \lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^{bn}; \quad \text{para} \quad a, b \in \mathbf{R}.$$

Así mismo, si $\lim_{n\to\infty} a_n = a \in \mathbf{R}$, entonces

$$e^a = \lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n . \tag{21}$$

Defina la función auxiliar $g_n(x) = (1 + x/n)^n$; para $1 + x/n > 0$ y $n \geq 1$. Entonces $g_n(x)$ es estrictamente creciente respecto de x. Considere $0 < \varepsilon < 1$. Salvo un número finito, todos los elementos de la sucesión $\{a_n\}_{n\geq 1}$ se encuentran en el intervalo $(a - \varepsilon, a + \varepsilon)$. Por lo que

$$e^{a-\varepsilon} = \lim_{n\to\infty} \left(1 + \frac{a-\varepsilon}{n}\right)^n \leq \lim_{n\to\infty} \left(1 + \frac{a_n}{n}\right)^n \leq \lim_{n\to\infty} \left(1 + \frac{a+\varepsilon}{n}\right)^n = e^{a+\varepsilon}.$$

Como ε es arbitrario, entonces se satisface (21).

Serie geométrica. Para $x \neq 0$ y $n \geq 0$, se tiene

$$1 + x + x^2 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1} = \frac{1 - x^{n+1}}{1 - x}.$$

De hecho

$$x^{n+1} - 1 = (x-1)(1 + x + \cdots + x^n); \quad \text{para} \quad x \in \mathbf{R}.$$

Por ejemplo

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} = 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1}{2}\right)^n = \frac{1 - (1/2)^{n+1}}{1 - 1/2} = 2\left(1 - \frac{1}{2^{n+1}}\right).$$

Al tomar el límite de esta expresión, se obtiene

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = 1 + \frac{1}{2} + \frac{1}{4} + \cdots = 2 \lim_{n\to\infty} \left(1 - \frac{1}{2^{n+1}}\right) = 2.$$

En general, para $|a| < 1$, se obtiene la serie conocida como *geométrica*:

$$\sum_{n=0}^{\infty} a^n = 1 + a + a^2 + \cdots = \frac{1}{1-a}. \tag{22}$$

E Teorema fundamental del Cálculo

Las siguientes son representaciones del teorema fundamental del cálculo.

1. Sea $f(x)$ una función integrable, definida en un intervalo cerrado $[a, b]$. Defina la función

$$F(x) = \int_a^x f(t)dt; \quad \text{para} \quad a \leq x \leq b.$$

   Entonces, la función $F(x)$ es derivable y

$$F'(x) = f(x); \quad \text{para} \quad a \leq x \leq b.$$

2. Sea $F(x)$ una función derivable. Entonces, $F'(x)$ es integrable y

$$\int_a^b F'(t)dt = F(b) - F(a); \quad \text{para} \quad a, b \in \mathbf{R}.$$

F Matrices y formas cuadráticas

Para $n \geq 1$, sea la matriz cuadrada de $n \times n$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}.$$

Esta matriz es invertible si y sólo si, su determinante es no nulo: $|\Sigma| \neq 0$. En tal caso

$$\Sigma\Sigma^{-1} = \Sigma^{-1}\Sigma = I_n,$$

donde Σ^{-1} denota la matriz *inversa* de Σ e I_n es la matriz *identidad* de $n \times n$. Si $x' = (x_1, \ldots, x_n) \in \mathbf{R}^n$, entonces

$$x'\Sigma x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij}. \tag{23}$$

De hecho

$$\begin{aligned}
x'\Sigma x &= (x_1, \ldots, x_n) \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= (x_1\sigma_{11} + \cdots + x_n\sigma_{n1}, \ldots, x_1\sigma_{1n} + \cdots + x_n\sigma_{nn}) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= x_1^2\sigma_{11} + \cdots + x_1x_n\sigma_{n1} + \cdots + x_1x_n\sigma_{1n} + \cdots + x_n^2\sigma_{nn} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} = \sum_{i=1}^{n} x_i^2 \sigma_{ii} + \sum_{i \neq j}^{n} x_i x_j \sigma_{ij}.
\end{aligned}$$

La matriz cuadrada Σ es *simétrica* si $\sigma_{ij} = \sigma_{ji}$; para $1 \leq i < j \leq n$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{pmatrix}.$$

En tal caso, a la expresión (23) se le dice *forma cuadrática*:

$$x'\Sigma x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \sigma_{ij} = \sum_{i=1}^{n} x_i^2 \sigma_{ii} + 2 \sum_{i<j}^{n} x_i x_j \sigma_{ij}. \tag{24}$$

La matriz simétrica Σ es *positiva definida* si

$$x'\Sigma x > 0; \quad \text{para todo} \quad x = (x_1, \ldots, x_n) \neq (0, \ldots, 0).$$

Aquí el vector x no es idénticamente cero en $\mathbf{R}^n$. Los *autovalores* (*eigenvalores*) de Σ son las soluciones reales o complejas de la ecuación

$$|\Sigma - tI_n| = 0; \quad t \in \mathbf{C}.$$

Hay a lo más n autovalores diferentes. En el caso de un autovalor complejo $t = u + iv$; con $u, v \in \mathbf{R}$ y $v \neq 0$ entonces, también es un autovalor su complejo conjugado $t^* = u - iv$.

Sea Σ una matriz simétrica de $n \times n$; con $n \geq 1$. Entonces, las siguientes afirmaciones son equivalentes

- Σ es positiva definida.

- $x'\Sigma x > 0$; para todo $x = (x_1, x_2) \neq (0,0)$.
- Los autovalores son positivos.

En tal caso

- $|\Sigma| > 0$.

Para el caso $n = 2$, el determinante de la matriz cuadrada

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

es

$$|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21} = ad - bc.$$

Recuerde que la matriz Σ invertible si y sólo si, su determinante es no nulo: $ad \neq bc$. En tal caso

$$\Sigma^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{25}$$

De hecho

$$\frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad - bc}\begin{pmatrix} ad - bc & bd - bd \\ -ac + ac & -bc + ad \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Si $x' = (x_1, x_2) \in \mathbf{R}^2$, entonces

$$x'\Sigma x = \sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j \sigma_{ij} = a x_1^2 + (b + c) x_1 x_2 + d x_2^2.$$

Si en particular Σ es simétrica, entonces

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}. \tag{26}$$

Su forma cuadrática queda como

$$x'\Sigma x = \sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j \sigma_{ij} = a x_1^2 + 2b x_1 x_2 + d x_2^2.$$

Por último, la matriz simétrica (26) es positiva definida si y sólo si, su determinante es positivo: $ad - b^2 > 0$.

G Desigualdades

Desigualdad de Cauchy Bunyakovsky Schwarz. Si $x = (x_1, \ldots, x_n)$ y $y = (x_1, \ldots, x_n)$, son vectores en $\mathbf{R}^n$, la desigualdad de Cauchy Bunyakovsky Schwarz es

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2},$$

que equivale a

$$\left(\sum_{i=1}^{n} x_i y_i \right)^2 \leq \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2. \tag{27}$$

Para verificar esta desigualdad, note que

$$\begin{aligned} |x|^2|y|^2 &= (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) \\ &= x_1^2 y_1^2 + \cdots + x_n^2 y_n^2 + x_1^2 y_2^2 + \cdots + x_1^2 y_n^2 + \cdots + x_n^2 y_1^2 + \cdots + x_n^2 y_{k-1}^2 \end{aligned}$$

y

$$\begin{aligned} |x \cdot y|^2 &= (x_1 y_1 + \cdots + x_n y_n)^2 \\ &= x_1^2 y_1^2 + \cdots + x_n^2 y_n^2 + 2(x_1 y_1)(x_2 y_2) + \cdots + 2(x_{n-1} y_{n-1})(x_n y_n). \end{aligned}$$

Entonces

$$\begin{aligned} & |x|^2|y|^2 - |x \cdot y|^2 \\ &= [x_1^2 y_2^2 - 2(x_1 y_2)(x_2 y_1) + x_2^2 y_1^2] + \cdots + [x_{n-1}^2 y_n^2 - 2(x_{n-1} y_n)(x_n y_{n-1}) + x_n^2 y_{n-1}^2] \\ &= (x_1 y_2 - x_2 y_1)^2 + \cdots + (x_{n-1} y_n - x_n y_{n-1})^2 \geq 0. \end{aligned}$$

Por otra parte, la versión integral de la desigualdad de Cauchy Bunyakovsky Schwarz es

$$\left(\int f(x) g(x) \mathrm{d}x \right)^2 \leq \int f^2(x) \mathrm{d}x \int g^2(x) \mathrm{d}x; \tag{28}$$

para todo par de funciones reales cuadrado integrables $f(x)$ y $g(x)$.

Si $x, y \in \mathbf{R}$ y $0 \leq r \leq s$, entonces

$$|x|^r |y|^{s-r} \leq |x|^s + |y|^s. \tag{29}$$

Si $|x| \leq |y|$, entonces

$$|x|^r |y|^{s-r} \leq |y|^r |y|^{s-r} = |y|^s \leq |x|^s + |y|^s.$$

En cambio, cuando $|y| \leq |x|$, se tiene

$$|x|^r |y|^{s-r} \leq |x|^r |x|^{s-r} = |x|^s \leq |x|^s + |y|^s.$$

Desigualdad de Young. Sean $u, v > 0$ y $0 < \alpha < 1$. Entonces

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1-\alpha)v. \tag{30}$$

La igualdad se cumple sólo para $u = v$. Para probar la desigualdad (30), divida esta expresión entre v:

$$\left(\frac{u}{v}\right)^\alpha \leq \alpha\left(\frac{u}{v}\right) + 1 - \alpha.$$

Sin pérdida de generalidad, considere $v = 1$. Defina la función auxiliar

$$g(u) = \alpha(u-1) - u^\alpha + 1; \quad \text{para} \quad u > 0.$$

Entonces

$$\begin{aligned} g'(u) &= \alpha - \alpha u^{\alpha-1} = \alpha(1 - u^{\alpha-1}), \\ g''(u) &= -\alpha(\alpha-1)u^{\alpha-2} = \alpha(1-\alpha)u^{\alpha-2} > 0. \end{aligned}$$

Por lo cual, la función $g(u)$ es estrictamente convexa y alcanza su mínimo global en $u = 1$, con $g(1) = 0$. Se concluye $g(u) > 0$; para $0 < u \neq 1$. Otra manera de probar la desigualdad de Young (30) es con la propiedad de concavidad de la función logaritmo:

$$\begin{aligned} u^\alpha v^{1-\alpha} &\leq \alpha u + (1-\alpha)v, \\ \alpha \log u + (1-\alpha)\log v &\leq \log(\alpha u + (1-\alpha)v). \end{aligned}$$

H Teorema de convergencia dominada.

Sea $\{X_n\}_{n\geq 0}$ una sucesión de variables aleatorias tal que $|X_n| \leq Y$; $n \geq 1$, para cierta variable aleatoria no negativa Y, con $\mathrm{E}\, Y < \infty$. Si $X_n \to X_0$, cuando $n \to \infty$ c.s., entonces, $\mathrm{E}\,|X_n| < \infty$; $n \geq 0$ y

$$\lim_{n\to\infty} \mathrm{E}\, X_n = \mathrm{E} \lim_{n\to\infty} X_n = \mathrm{E}\, X_0 \tag{31}$$

I Resumen de propiedades de distribuciones de probabilidad

Las siguientes dos tablas resumen las propiedades de algunas distribuciones de probabilidad paramétricas, tanto discretas como continuas.

Tabla 1. Propiedades de algunas distribuciones de probabilidad discretas.

distribución	función de densidad $f(x)$	espacio paramétrico	moda y mediana	esperanza y varianza	F.G.M. $M(t) = \mathrm{E}\, e^{tX}$
uniforme discreta $\mathrm{U}\{a, \ldots, b\}$	$\dfrac{1}{b-a+1}$ $x = a, a+1, \ldots, b$	$a \leq b$, $a, b \in \mathbf{Z}$	$a, \ldots, b$ $\dfrac{a+b}{2}$	$(a+b)/2$ $\dfrac{(b-a+1)^2 - 1}{12}$	$\dfrac{\sum_{x=a}^{b} e^{tx}}{b-a+1}$
Bernoulli $\mathrm{Ber}(p)$	$p^x(1-p)^{1-x}$ $x = 0, 1$	$0 \leq p \leq 1$	$1_{\{p \geq 1/2\}}$ $1_{\{p > 1/2\}}$	p $p(1-p)$	$pe^t + 1 - p$
binomial $\mathrm{bin}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$0 \leq p \leq 1$	$[(n+1)p] - 1$ $[(n+1)p]$ $[np]$ ó $[np] + 1$	np $np(1-p)$	$(pe^t + 1 - p)^n$
Poisson $\mathrm{Po}(\mu)$	$e^{-\mu} \dfrac{\mu^x}{x!}$ $x = 0, 1, \ldots$	$\mu > 0$	no explícita $[\mu] - 1$ ó $[\mu]$	μ μ	$e^{-\mu(e^t - 1)}$
geométrica $\mathrm{geo}(p)$	$p(1-p)^x$ $x = 0, 1, \ldots$	$0 < p < 1$	0 $\left\lceil \dfrac{-1}{\log_2(1-p)} \right\rceil - 1$	$(1-p)/p$ $\dfrac{1-p}{p^2}$	$\dfrac{p}{1-(1-p)e^t}$ $t < -\log(1-p)$
geométrica $\mathrm{geo}(p)$	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$0 < p < 1$	1 $\left\lceil \dfrac{-1}{\log_2(1-p)} \right\rceil$	$1/p$ $\dfrac{1-p}{p^2}$	$\dfrac{pe^t}{1-(1-p)e^t}$ $t < -\log(1-p)$
binomial negativa $\mathrm{binneg}(r, p)$	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, \ldots$	$0 < p < 1$	$r + \left[\dfrac{(r-1)(1-p)}{p} \right]$ no explícita	r/p $\dfrac{r(1-p)}{p^2}$	$\left(\dfrac{pe^t}{1-(1-p)e^t} \right)^r$ $t < -\log(1-p)$

Notas: $[x]$ y $\lceil x \rceil$ denotan respectivamente los enteros más cercanos a la izquierda y a la derecha de x; para $x \geq 0$.

Tabla 2. Propiedades de algunas distribuciones de probabilidad continuas.

distribución	función de densidad $f(x)$	espacio paramétrico	moda y mediana	esperanza y varianza	F.G.M. $M(t) = \mathrm{E}\, e^{tX}$
uniforme continua $\mathrm{U}(a,b)$	$\dfrac{1}{b-a}$ $a < x < b$	$a < b$	(a,b) $\dfrac{a+b}{2}$	$(a+b)/2$ $\dfrac{(b-a)^2}{12}$	$\dfrac{e^{bt}-e^{at}}{(b-a)t}$
$\text{beta}(\alpha,\beta)$	$\dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(m,n)}$ $0 < x < 1$	$\alpha > 0$ $\beta > 0$	$\dfrac{\alpha-1}{\alpha+\beta-2}$; $\alpha, \beta > 1$ no explícita	$\dfrac{\alpha}{\alpha+\beta}$ $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	no explícita
exponencial $\exp(\lambda)$	$\lambda e^{-\lambda x}$ $x > 0$	$\lambda > 0$ $\mu = \dfrac{1}{\lambda}$	0 $\dfrac{\log 2}{\lambda}$	$1/\lambda$ $1/\lambda^2$	$\dfrac{\lambda}{\lambda - t}$ $t < \lambda$
$\text{Erlang}(n,\lambda)$	$\dfrac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}$ $x > 0$	$\lambda > 0$ $\beta = 1/\lambda$ $n = 1, 2, \ldots$	$(n-1)\lambda$ no explícita	n/λ n/λ^2	$\left(\dfrac{\lambda}{\lambda - t}\right)^n$ $t < \lambda$
χ-cuadrada χ^2_n	$\dfrac{x^{n/2-1}}{2^{n/2}\,\Gamma(n/2)} e^{-x/2}$ $x > 0$	$n > 0$	$0 \vee (n-2)$ no explícita	n $2n$	$\dfrac{1}{(1-2t)^{n/2}}$ $t < 1/2$
$\text{gamma}(\alpha,\lambda)$	$\dfrac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}$ $x > 0$	$\alpha > 0$ $\lambda > 0$ $\beta = 1/\lambda$	$\dfrac{0 \vee (\alpha-1)}{\lambda}$ no explícita	α/λ α/λ^2	$\left(\dfrac{\lambda}{\lambda - t}\right)^\alpha$ $t < \lambda$
$\text{Weibull}(\alpha,\lambda)$	$\alpha\lambda^\alpha x^{\alpha-1} e^{-(\lambda x)^\alpha}$ $x > 0$	$\alpha > 0$ $\lambda > 0$ $\beta = \dfrac{1}{\lambda}$	$\dfrac{\left(0 \vee \frac{\alpha-1}{\alpha}\right)^{1/\alpha}}{\lambda}$ $\dfrac{(\log 2)^{1/\alpha}}{\lambda}$	$\dfrac{\Gamma(1+1/\alpha)}{\lambda}$ $\dfrac{\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})}{\lambda^2}$	no explícita

distribución	función de densidad $f(x)$	espacio paramétrico	moda y mediana	esperanza y varianza	F.G.M. $M(t) = \mathrm{E}\, e^{tX}$
normal $\mathrm{N}(\mu, \sigma^2)$	$\dfrac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $x \in \mathbf{R}$	$\mu \in \mathbf{R}$ $\sigma^2 > 0$	μ μ	μ σ^2	$e^{\mu t + \sigma^2 t^2/2}$
log normal $\log\mathrm{N}(\mu, \sigma^2)$	$\dfrac{1}{\sqrt{2\pi}\sigma x}\, e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}$ $x > 0$	$\mu \in \mathbf{R}$ $\sigma^2 > 0$	$e^{\mu-\sigma^2}$ e^{μ}	$e^{\mu+\sigma^2/2}$ $e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$	finita para $t \leq 0$
Student t_n	$\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{(n+1)/2}}$ $x \in \mathbf{R}$	$n > 0$	0 0	$0;\ n > 1$ $\dfrac{n}{n-2};\ n > 2$	no definida
Cauchy(μ, σ)	$\dfrac{1}{\pi\sigma}\dfrac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2}$ $x \in \mathbf{R}$	$\mu \in \mathbf{R}$ $\sigma > 0$	μ μ	no definidos	no definida
Fisher $\mathrm{F}_{m,n}$	$\dfrac{\left(\frac{m}{n}\right)^{m/2} x^{m/2-1}}{\mathrm{B}\left(\frac{m}{2}, \frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{(m+n)/2}}$ $x > 0$	$m > 0$ $n > 0$	$\dfrac{0 \vee (m-2)n}{m(n+2)}$ no explícita	$\dfrac{n}{n-2};\ n > 2$ $\dfrac{2n^2(m+n-2)}{m(n-2)(n-4)};$ $n > 4$	finita para $t \leq 0$

Notas: las funciones *gamma* y *beta* son respectivamente $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x$ y $\mathrm{B}(\alpha,\beta) = \Gamma(\alpha)\,\Gamma(\beta)/\,\Gamma(\alpha+\beta)$, para $\alpha, \beta > 0$.

Índice analítico

Referencias

[1] Tom M. Apostol. *Mathematical Analysis*. 2nd. Reading: Addison-Wesley Publishing Company, Inc., 1974.

[2] William Feller. *An introduction to probability theory and its applications*. 2nd. New York: John Wiley y Sons, 1971.

[3] Louis Gordon. “A Stochastic approach to the Gamma function”. En: *The American Mathematical Monthly* 101.9 (1994), págs. 858-865. DOI: doi.org/10.1080/00029890.1994.11997039.

[4] IMSS. *Puestos de trabajo afiliados al Instituto Mexicano del Seguro Social, Boletín de prensa núm. 005/2024*. 2024. URL: https://www.imss.gob.mx/sites/all/statics/i2f_news/005.-Bolet%C3%ADn%20Empleo%20Diciembre%202023.pdf (visitado 28-05-2024).

[5] Reporte Índigo. *La Ouija del Diablo*. 2015. URL: https://www.reporteindigo.com/reporte/la-ouija-del-diablo (visitado 16-08-2023).

[6] INEGI. *Encuesta nacional de ingresos y gastos de los hogares (ENIGH) 2022, Comunicado de prensa núm. 420/23*. URL: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2023/ENIGH/ENIGH2022.pdf (visitado 26-07-2023).

[7] INEGI. *Índice Nacional de Precios al Consumidor (INPC) 2024, Comunicado de prensa núm. 8/24*. 2024. URL: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/inpc_2q/inpc_2q2024_01.pdf (visitado 28-05-2024).

[8] INEGI. *Vehículos de motor registrados en circulación, INEGI*. URL: https://www.inegi.org.mx/temas/vehiculos (visitado 28-06-2023).

[9] Douglas C. Montgomery y George C. Runger. *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.

[10] Miguel Nakamura y Victor Pérez-Abreu. “Use of an empirical probability generating function for testing a Poisson model”. En: *Canadian Journal of Statistics* 21.2 (1993), págs. 149-156. DOI: doi.org/10.2307/3315808.

[11] H.L. Royden y P. Fitzpatrick. *Real Analysis*. 4th. China Machine Press, 2010.

[12] Walter Rudin. *Principles of mathematical analysis*. 3rd. McGraw-Hill, 1976.

[13] Michael Spivac. *Cálculo infinitesimal*. 2da. Barcelona: Reverté, 1992.

[14] R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023. URL: https://www.R-project.org (visitado 15-08-2023).

[15] Wikipedia. *Detector de mentiras — Wikipedia, La enciclopedia libre*. 2023. URL: https://es.wikipedia.org/w/index.php?title=Detector_de_mentiras&oldid=152530896 (visitado 15-08-2023).

[16] Cheh-Chih Yeh, Hung-Wen Yeh y Wenyaw Chan. "Some equivalent forms of the arithematic-geometric mean inequality in probability: A survey". En: *Journal of Inequalities and Applications* 2008.1 (2008), pág. 386715. DOI: doi.org/10.1155/2008/386715.