

# Introducción a la probabilidad

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

Agosto 2013  
1era. revisión: diciembre 2013

## Prólogo

El presente trabajo es un ensayo que contiene material suficiente para un primer curso a nivel universitario sobre la teoría de la probabilidad. Está dirigido a estudiantes de las carreras de matemáticas, actuaría, matemáticas aplicadas y otras carreras científicas similares cuyos programas de estudio contemplan un curso semestral en donde se muestren los resultados, usos y aplicaciones de la probabilidad elemental. Se estudian temas tradicionales de la probabilidad básica, las variables aleatorias más conocidas y sus distribuciones de probabilidad, así como algunas técnicas y resultados clásicos de la probabilidad. Se ha buscado que en el texto aparezcan numerosas gráficas y diagramas con el objetivo de hacer las explicaciones más claras. También se han incluido sugerencias del uso del paquete computacional R para el estudio y simulación de las variables aleatorias y sus distribuciones. En particular, a través de unas pocas instrucciones en R se muestra la forma en la que se puede corroborar experimentalmente tanto la ley de los grandes números como el teorema central del límite. El texto fue escrito en el sistema  $\text{\LaTeX}$  y las ilustraciones fueron elaboradas usando el paquete PSTricks. Para una lectura provechosa de este material, se requiere tener cierta familiaridad con algunos conceptos del álgebra y del cálculo diferencial e integral.

Agradezco sinceramente a los revisores anónimos quienes emitieron dictámenes constructivos acerca de este trabajo. Agradezco también el apoyo financiero del programa PAPIME Proyecto PE101113, DGAPA, UNAM, a través del cual fue posible la impresión de este texto y la elaboración del curso en videos disponible en la página web

<http://www.matematicas.unam.mx/lars/0625>

Luis Rincón  
Agosto 2013  
Ciudad Universitaria UNAM  
`lars@ciencias.unam.mx`

# Contenido

<b>1. Probabilidad elemental</b>	<b>1</b>
1.1. Experimentos aleatorios . . . . .	3
1.2. Espacio muestral . . . . .	5
1.3. Operaciones con conjuntos . . . . .	8
1.4. Probabilidad clásica . . . . .	18
1.5. Probabilidad geométrica . . . . .	21
1.6. Probabilidad frecuentista . . . . .	28
1.7. Probabilidad subjetiva . . . . .	31
1.8. Probabilidad axiomática . . . . .	32
1.9. Sigmas álgebras . . . . .	42
1.10. Sigma álgebra de Borel . . . . .	45
1.11. Espacios de probabilidad . . . . .	46
1.12. Análisis combinatorio . . . . .	49
1.13. Probabilidad condicional . . . . .	63
1.14. Teorema de probabilidad total . . . . .	68
1.15. Teorema de Bayes . . . . .	73
1.16. Independencia de eventos . . . . .	79
1.17. Continuidad de las medidas de probabilidad . . . . .	87
<b>2. Variables aleatorias</b>	<b>95</b>
2.1. Variables aleatorias . . . . .	95
2.2. Función de probabilidad . . . . .	103
2.3. Función de distribución . . . . .	112
2.4. Teorema de cambio de variable . . . . .	123
2.5. Independencia de variables aleatorias . . . . .	126
2.6. Esperanza . . . . .	130

---

2.7. Varianza . . . . .	142
2.8. Momentos . . . . .	149
2.9. Cuantiles . . . . .	152
2.10. Función generadora de probabilidad . . . . .	155
2.11. Función generadora de momentos . . . . .	160
<b>3. Distribuciones de probabilidad</b> . . . . .	<b>165</b>
3.1. Distribución uniforme discreta . . . . .	165
3.2. Distribución Bernoulli . . . . .	169
3.3. Distribución binomial . . . . .	174
3.4. Distribución geométrica . . . . .	181
3.5. Distribución binomial negativa . . . . .	185
3.6. Distribución hipergeométrica . . . . .	189
3.7. Distribución Poisson . . . . .	193
3.8. Distribución uniforme continua . . . . .	202
3.9. Distribución exponencial . . . . .	206
3.10. Distribución gama . . . . .	210
3.11. Distribución beta . . . . .	215
3.12. Distribución Weibull . . . . .	218
3.13. Distribución normal . . . . .	221
3.14. Distribución ji-cuadrada . . . . .	229
3.15. Distribución $t$ . . . . .	233
3.16. Distribución $F$ . . . . .	237
<b>4. Vectores aleatorios</b> . . . . .	<b>241</b>
4.1. Vectores aleatorios . . . . .	241
4.2. Función de probabilidad conjunta . . . . .	243
4.3. Función de distribución conjunta . . . . .	253
4.4. Función de probabilidad marginal . . . . .	258
4.5. Función de distribución marginal . . . . .	262
4.6. Independencia de variables aleatorias . . . . .	264
4.7. Distribución condicional . . . . .	268
4.8. Esperanza condicional . . . . .	272
4.9. Covarianza . . . . .	274
4.10. Coeficiente de correlación . . . . .	277
<b>5. Teoremas límite</b> . . . . .	<b>281</b>

---

5.1. Desigualdad de Chebyshev . . . . .	281
5.2. Convergencia de variables aleatorias . . . . .	284
5.3. La ley de los grandes números . . . . .	288
5.4. El teorema central del límite . . . . .	295
<b>A. Apéndice</b>	<b>309</b>
A.1. Notación . . . . .	309
A.2. El alfabeto griego . . . . .	309
A.3. Exponentes y logaritmos . . . . .	310
A.4. Fórmulas para sumas . . . . .	311
A.5. Fórmulas de derivación e integración . . . . .	312
A.6. El lema de Abel . . . . .	313
A.7. Notación $o$ -pequeña . . . . .	313
A.8. Tabla de la distribución normal estándar . . . . .	315
A.9. Tabla de la distribución $t(n)$ . . . . .	316
A.10. Tabla de la distribución $\chi^2(n)$ . . . . .	317
<b>Índice analítico</b>	<b>321</b>



# Capítulo 1

## Probabilidad elemental

En esta primera parte estudiaremos algunos de los conceptos más elementales de la teoría de la probabilidad. Esta teoría matemática tuvo como uno de sus primeros puntos de partida el intentar resolver un problema particular concerniente a una apuesta de juego de dados entre dos personas. El problema al que nos referimos involucraba una gran cantidad de dinero y puede plantearse de la siguiente forma:

*Dos jugadores escogen cada uno de ellos un número del 1 al 6, distinto uno del otro, y apuestan 32 doblones de oro a que el número escogido por uno de ellos aparece en tres ocasiones antes que el número del contrario al lanzar sucesivamente un dado. Suponga que el número de uno de los jugadores ha aparecido dos veces y el número del otro una sola vez. Bajo estas circunstancias, ¿cómo debe dividirse el total de la apuesta si el juego se suspende?*

Uno de los apostadores, Antonio de Gombaud, popularmente conocido como el caballero De Méré, deseando conocer la respuesta al problema plantea la situación a Blaise Pascal (1623-1662). Pascal a su vez consulta con Pierre de Fermat (1601-1665) e inician estos últimos un intercambio de cartas a propósito del problema. Esto sucede en el año de 1654. Con ello se inician algunos esfuerzos por dar solución a éste y otros problemas similares que se plantean. Con el paso del tiempo se sientan las bases y las experiencias necesarias para la búsqueda de una teoría matemática que sintetice los

conceptos y los métodos de solución de los muchos problemas particulares resueltos a lo largo de varios años. En el segundo congreso internacional de matemáticas, celebrado en la ciudad de París en el año 1900, el matemático David Hilbert (1862-1943) plantea 23 problemas matemáticos de importancia. Uno de estos problemas es el de encontrar axiomas o postulados a partir de los cuales se pueda construir una teoría matemática de la probabilidad. Aproximadamente treinta años después, en 1933, el matemático ruso A. N. Kolmogorov (1903-1987) propone ciertos axiomas que a la postre resultaron adecuados para la construcción de una teoría de la probabilidad. Esta teoría prevalece hoy en día y ha adquirido el calificativo de [teoría clásica](#).



Blaise Pascal  
(Francia 1623–1662)



Pierre de Fermat  
(Francia 1601–1665)

Actualmente la teoría de la probabilidad se ha desarrollado y extendido enormemente gracias a muchos pensadores que han contribuido a su crecimiento, y es sin duda una parte muy importante y bien establecida de las matemáticas. La teoría de la probabilidad ha resultado muy útil para modelar fenómenos de muy diversas disciplinas del conocimiento humano en donde es necesario incorporar la incertidumbre o el azar como un elemento del modelo.

## 1.1. Experimentos aleatorios

Existen dos tipos de fenómenos o experimentos en la naturaleza: los deterministas y los aleatorios. Un **experimento determinista** es aquel que produce el mismo resultado cuando se le repite bajo las mismas condiciones, por ejemplo, medir el volumen de un gas cuando la presión y la temperatura son constantes produce teóricamente siempre el mismo resultado, o medir el ángulo de un rayo de luz reflejado en un espejo resulta siempre en el mismo resultado cuando el ángulo de incidencia es el mismo y el resto de las condiciones son constantes. Muchas otras leyes de la física son ejemplos de situaciones en donde bajo idénticas condiciones iniciales, el resultado del experimento es siempre el mismo. En contraparte, un **experimento aleatorio** es aquel que cuando se le repite bajo las mismas condiciones, el resultado que se observa no siempre es el mismo y tampoco es predecible. El lanzar una moneda al aire y observar la cara de la moneda que mira hacia arriba, o registrar el número ganador en un juego de lotería, son ejemplos cotidianos de experimentos aleatorios.

Nuestro interés consiste en estudiar algunos modelos matemáticos que se han logrado desarrollar a lo largo de varios años con el objetivo de contar con el conocimiento de conceptos y resultados que nos permitan tener un mejor entendimiento y control de los muy diversos fenómenos aleatorios presentes en la vida del hombre.

Para ser más precisos, pediremos que los experimentos aleatorios que consideraremos cumplan las siguientes características:

- a) El experimento debe poder ser repetible bajo las mismas condiciones iniciales.
- b) El resultado de cualquier ensayo del experimento es variable y depende del azar o de algún mecanismo aleatorio.

Es necesario mencionar, sin embargo, que en algunas ocasiones no es evidente poder clasificar un experimento dado en aleatorio o determinista, esto dependerá del observador, de lo que él o ella conozca del experimento y de lo que esta persona desea observar en el experimento. Así, el experimento

mismo no está separado completamente del observador, pues la concepción y entendimiento del experimento dependen en alguna medida del observador. En la siguiente sección de ejercicios se muestran algunos casos particulares.

## Ejercicios

1. Clasifique los siguientes experimentos en deterministas o aleatorios:
  - a) Registrar el número de accidentes que ocurren en una determinada calle de una ciudad.
  - b) Observar la temperatura a la que hierve el agua a una altitud dada.
  - c) Registrar el consumo de electricidad de una casa-habitación en un día determinado.
  - d) Registrar la hora a la que desaparece el sol en el horizonte en un día dado visto desde una posición geográfica determinada.
  - e) Observar el precio que tendrá el petróleo en un año.
  - f) Registrar la altura máxima que alcanza un proyectil lanzado verticalmente.
  - g) Observar el número de años que vivirá un bebé que nace en este momento.
  - h) Observar el ángulo de reflexión de un haz de luz incidente en un espejo.
  - i) Registrar la precipitación pluvial anual en una zona geográfica determinada.
  - j) Observar el tiempo que tarda un objeto en caer al suelo cuando se le deja caer desde una altura dada.
  - k) Registrar el ganador de una elección en un proceso de votación libre y secreto.
  - l) Observar la posición de una molécula de oxígeno en una habitación después de dejarla en libre movimiento durante un minuto.

## 1.2. Espacio muestral

Hemos mencionado que la teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios. En principio no sabemos cuál será el resultado de un experimento aleatorio, así que por lo menos conviene agrupar en un conjunto a todos los resultados posibles. Esto lleva a la siguiente definición.

**Definición 1.1** El **espacio muestral**, o también llamado **espacio muestra**, de un experimento aleatorio es el conjunto de todos los posibles resultados del experimento, y se le denota generalmente por la letra griega  $\Omega$  (omega mayúscula). A un resultado particular se le denota por  $\omega$  (omega minúscula).

Más adelante mostraremos que el espacio muestral no es necesariamente único y su determinación depende del interés o la necesidad de la persona que realiza el experimento aleatorio. En algunos textos se usa también la letra  $S$  para denotar al espacio muestral. Esta letra proviene del término *Sampling space* de la lengua inglesa, equivalente a espacio muestral. Por otro lado y de manera preliminar llamaremos evento o suceso a cualquier subconjunto del espacio muestral y los denotaremos por las primeras letras del alfabeto en mayúsculas:  $A, B, C, \dots$  o bien por alguna otra letra en mayúscula que nos ayude a identificar al evento. A través de algunos ejemplos ilustraremos a continuación los conceptos de espacio muestral y evento.

**Ejemplo 1.1** Si un experimento aleatorio consiste en lanzar un dado y observar el número que aparece en la cara superior, entonces claramente el espacio muestral es el conjunto  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Como ejemplo de un evento para este experimento podemos definir el conjunto  $A = \{2, 4, 6\}$ , que corresponde al suceso de obtener como resultado un número par. Si al lanzar el dado una vez se obtiene el número “4”, decimos entonces que se observó la ocurrencia del evento  $A$ , y si se obtiene por ejemplo el resultado “1”, decimos que no se observó la ocurrencia del evento  $A$ . •

**Ejemplo 1.2** Considere el experimento aleatorio de participar en un juego de lotería. Suponga que hay un millón de números en esta lotería y un jugador participa con un boleto. ¿Cuál es un posible espacio muestral para este experimento si únicamente uno de los posibles números es el ganador? Naturalmente al jugador le interesa conocer su suerte en este juego y puede proponer como espacio muestral el conjunto  $\Omega = \{\text{“ganar”}, \text{“perder”}\}$ . Sin embargo puede también tomarse como espacio muestral el conjunto que contiene a todos los posibles números ganadores, es decir,  $\Omega = \{1, 2, \dots, 1000000\}$ . Este ejemplo sencillo muestra que el espacio muestral de un experimento aleatorio no es único y depende del interés del observador. •

**Ejemplo 1.3** Suponga que un experimento aleatorio consiste en observar el tiempo en el que una máquina en operación sufre su primera descompostura. Si se consideran mediciones continuas del tiempo, entonces puede adoptarse como espacio muestral para este experimento el intervalo  $[0, \infty)$ . El subconjunto  $A = [1, 2]$  corresponde al evento en el que la primera descompostura se observe entre la primera y la segunda unidad de tiempo. •

Se dice que un evento es **simple** cuando consta de un solo elemento del espacio muestral, en cambio, se llama **compuesto** cuando consta de más de un elemento del espacio muestral. Puesto que los conceptos de espacio muestral y evento involucran forzosamente el manejo de conjuntos, recordaremos en la siguiente sección algunas operaciones entre estos objetos y algunas propiedades que nos serán de utilidad en el estudio de la probabilidad.

## Ejercicios

2. Determine un espacio muestral para el experimento aleatorio de:
  - a) Observar la posición de un partícula en un instante dado, la cual se mueve sin restricciones en un espacio tridimensional.
  - b) Registrar el número de personas que requieren hospitalización en el siguiente accidente automovilístico atendido por los servicios de emergencia en una localidad dada.

- 
- c)* Lanzar un dado hasta que se obtiene un “6”.
  - d)* Registrar la fecha de cumpleaños de  $n$  personas escogidas al azar.
  - e)* Observar la forma en la que  $r$  personas que abordan un elevador en la planta baja de un edificio descienden en los pisos  $1, 2, \dots, n$ .
  - f)* Registrar la duración de una llamada telefónica escogida al azar.
  - g)* Observar el número de años que le restan de vida a una persona escogida al azar dentro del conjunto de asegurados de una compañía aseguradora.
3. Proponga un espacio muestral para el experimento aleatorio de lanzar tres monedas a un mismo tiempo suponiendo que las monedas:
- a)* Son distinguibles.
  - b)* No son distinguibles.
4. Considere el experimento aleatorio de lanzar dos dados distinguibles. Escriba explícitamente los resultados asociados a los siguientes eventos y determine su cardinalidad.
- a)*  $A =$  “La suma de los dos resultados es 7.”
  - b)*  $B =$  “Uno de los dos dados cae en número impar y el otro en número par.”
  - c)*  $C =$  “El resultado de un dado difiere del otro en a lo sumo una unidad.”
  - d)*  $D =$  “El resultado de un dado difiere del otro en por lo menos cuatro unidades.”
  - e)*  $E = A \cap B$ .
  - f)*  $F = B^c$ .
  - g)*  $G = C \cup D$ .

### 1.3. Operaciones con conjuntos

Supondremos que el espacio muestral  $\Omega$  de un experimento aleatorio es una especie de conjunto universal y cualquier elemento de  $\Omega$  lo denotaremos por  $\omega$  (omega minúscula). Al conjunto vacío lo denotaremos como es usual por el símbolo  $\emptyset$ . Otros símbolos usuales son los de pertenencia ( $\in$ ), o no pertenencia ( $\notin$ ) de un elemento en un conjunto, y los de contención ( $\subset, \subseteq$ ), o no contención ( $\not\subset$ ) de un conjunto en otro. ¿Puede usted explicar el significado del símbolo  $\subsetneq$ ? Justamente, decimos que  $A$  es un subconjunto propio de  $B$  si  $A \subsetneq B$ . La igualdad de dos conjuntos  $A$  y  $B$  significa que se cumplen las dos contenciones:  $A \subset B$  y  $B \subset A$ . Por último, si  $A$  es un conjunto, denotamos la cardinalidad o número de elementos de ese conjunto por el símbolo  $\#A$ .

Sean  $A$  y  $B$  dos subconjuntos cualesquiera de  $\Omega$ . Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento:

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ ó } \omega \in B\},$$

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ y } \omega \in B\},$$

$$A - B = \{\omega \in \Omega : \omega \in A \text{ y } \omega \notin B\},$$

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

Cuando los conjuntos se expresan en palabras, la operación unión,  $A \cup B$ , se lee “A ó B” y la intersección,  $A \cap B$ , se lee “A y B”. En la Figura 1.1 se muestran en diagramas de Venn estas dos operaciones.

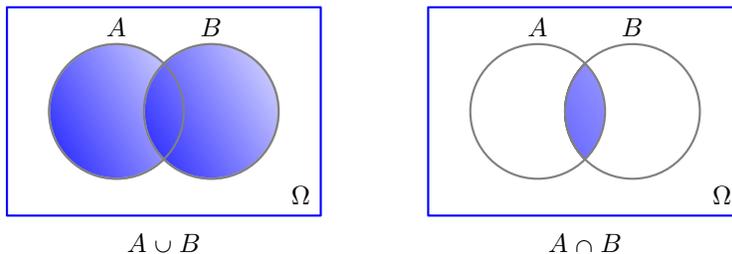


Figura 1.1

La diferencia entre dos conjuntos  $A$  y  $B$  se denota por  $A - B$ , y corresponde a aquel conjunto de elementos de  $A$  que no pertenecen a  $B$ , es decir,  $A - B$

se define como  $A \cap B^c$ . En general, el conjunto  $A - B$  es distinto de  $B - A$ , de hecho estos conjuntos son siempre ajenos, ¿puede usted comprobar tal afirmación? ¿en qué caso ambos conjuntos coinciden? Por otro lado el complemento de un conjunto  $A$  se denota por  $A^c$  y se define como la colección de aquellos elementos de  $\Omega$  que no pertenecen a  $A$ . Mediante un diagrama de Venn ilustramos gráficamente las operaciones de diferencia y complemento en la Figura 1.2.

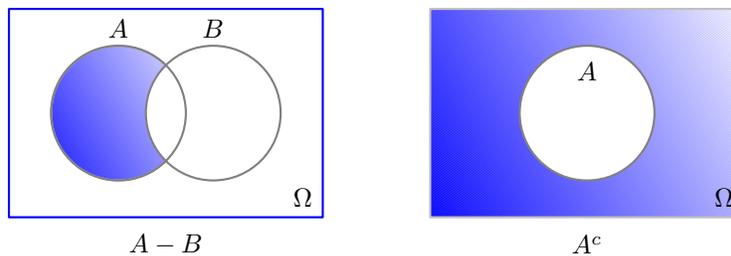


Figura 1.2

**Ejemplo 1.4** Sea  $A$  el conjunto de aquellas personas que tienen hijos y  $B$  la colección de aquellas personas que están casadas. Entonces el conjunto  $A \cap B$  consta de aquellas personas que están casadas y tienen hijos, mientras que el conjunto  $A \cap B^c$  está constituido por aquellas personas que tienen hijos pero no están casadas. ¿Quién es  $A^c \cap B$ ? Observe que cada persona es un elemento de alguno de los siguientes conjuntos:  $A \cap B$ ,  $A \cap B^c$ ,  $A^c \cap B$  ó  $A^c \cap B^c$ . ¿A cuál pertenece usted? •

Es fácil verificar que el conjunto vacío y el conjunto total satisfacen las siguientes propiedades elementales:  $A \cup \emptyset = A$ ,  $A \cap \Omega = A$ ,  $A \cap \emptyset = \emptyset$ ,  $A \cup A^c = \Omega$ ,  $A \cup \Omega = \Omega$ ,  $A \cap A^c = \emptyset$ . Además las operaciones unión e intersección son asociativas, esto es, satisfacen las siguientes igualdades:

$$A \cup (B \cap C) = (A \cup B) \cap C,$$

$$A \cap (B \cup C) = (A \cap B) \cup C,$$

y también son distributivas, es decir,

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

Recordemos también la operación diferencia simétrica entre dos conjuntos  $A$  y  $B$ , denotada por  $A \Delta B$ , y definida como sigue

$$A \Delta B = (A \cup B) - (B \cap A).$$

En la Figura 1.3 ilustramos gráficamente el conjunto resultante de efectuar la diferencia simétrica entre los conjuntos  $A$  y  $B$ . Visualmente es fácil comprobar que la diferencia simétrica también puede escribirse como  $(A - B) \cup (B - A)$ . ¿Cómo podría expresarse en palabras al conjunto  $A \Delta B$ ?

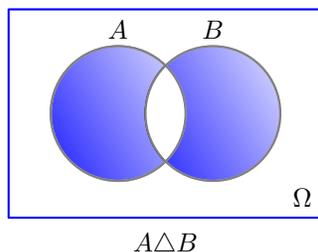


Figura 1.3

Recordemos además las muy útiles leyes de De Morgan:

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

La validez de estas dos igualdades puede extenderse a colecciones finitas e incluso arbitrarias de conjuntos. ¿Puede usted escribir estas identidades para  $n$  conjuntos?

### Conjuntos ajenos

Cuando dos conjuntos no tienen ningún elemento en común se dice que son ajenos, es decir, los conjuntos  $A$  y  $B$  son ajenos o disjuntos si se cumple la igualdad

$$A \cap B = \emptyset.$$

Por ejemplo, si  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , entonces los conjuntos  $A = \{1, 2\}$  y  $B = \{3, 4\}$  son ajenos pues no hay ningún elemento común entre ellos. El ejemplo general más importante de conjuntos o eventos ajenos es la pareja dada por  $A$  y  $A^c$ , para cualquier conjunto  $A$ . La propiedad de ser ajenos puede extenderse al caso cuando se tienen varios conjuntos: Decimos que  $n$  conjuntos  $A_1, \dots, A_n$  son ajenos si  $A_1 \cap \dots \cap A_n = \emptyset$ , y se dice que son ajenos dos a dos (o mutuamente ajenos) si  $A_i \cap A_j = \emptyset$  para cualesquiera valores de los índices  $i, j = 1, 2, \dots, n$ , con  $i$  distinto de  $j$ . La propiedad de ser ajenos dos a dos para una colección de eventos implica que los conjuntos son ajenos, sin embargo, el hecho de que todos ellos sean ajenos no implica que sean ajenos dos a dos. Es decir, la propiedad de ser ajenos dos a dos es más fuerte que la propiedad de ser simplemente ajenos. Ilustraremos la situación con el siguiente ejemplo.

**Ejemplo 1.5** Los conjuntos  $A = \{1, 2\}$ ,  $B = \{2, 3\}$  y  $C = \{3, 4\}$  son ajenos pues  $A \cap B \cap C = \emptyset$ , pero no son ajenos dos a dos pues, por ejemplo, el conjunto  $A \cap B$  no es vacío. Así, los conjuntos  $A$ ,  $B$  y  $C$  son ajenos en el sentido de que la intersección de todos ellos es vacía pero no son ajenos dos a dos. •

Las operaciones entre conjuntos que mencionaremos a continuación no son elementales y producen nuevos conjuntos que se encuentran en un nivel distinto al de los conjuntos originales.

### Conjunto potencia

El conjunto potencia de  $\Omega$ , denotado por  $2^\Omega$ , es aquel conjunto constituido por todos los subconjuntos posibles de  $\Omega$ . En términos estrictos esta nueva colección deja de ser un conjunto y se le llama clase de subconjuntos de  $\Omega$ , aunque seguiremos usando el primer término en nuestro tratamiento elemental de conjuntos. Por ejemplo, si  $\Omega = \{a, b, c\}$ , entonces el conjunto  $2^\Omega$  consta de 8 elementos, a saber,

$$2^\Omega = \left\{ \emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \Omega \right\}.$$

Observe que los elementos del conjunto potencia son en sí mismos conjuntos, y que en esta colección están contenidos todos los eventos que podrían ser

de interés en un experimento aleatorio. No es difícil demostrar que cuando  $\#\Omega < \infty$ ,

$$\#(2^\Omega) = 2^{\#\Omega},$$

es decir, el número de elementos en el conjunto  $2^\Omega$  es exactamente 2 elevado a la potencia dada por la cardinalidad de  $\Omega$ . De este hecho proviene la notación usada para el conjunto potencia:  $2^\Omega$ . Observe que la expresión  $2^\Omega$  no tiene el significado matemático del número 2 elevado a la potencia  $\Omega$  pues ello no tiene sentido. Debe considerarse como un símbolo para denotar al conjunto potencia y que sugiere el número de elementos en dicha clase. Para el ejemplo anterior se comprueba que la cardinalidad de  $2^\Omega$  es efectivamente  $2^{\#\Omega} = 2^3 = 8$ .

### Producto Cartesiano

El producto Cartesiano de dos conjuntos  $A$  y  $B$ , denotado por  $A \times B$ , se define como la colección de todas las parejas ordenadas  $(a, b)$ , en donde  $a$  es cualquier elemento de  $A$ , y  $b$  es cualquier elemento de  $B$ . En símbolos,

$$A \times B = \{(a, b) : a \in A \text{ y } b \in B\}.$$

**Ejemplo 1.6** Si  $A = \{a_1, a_2\}$  y  $B = \{b_1, b_2, b_3\}$ , entonces

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_1), (a_2, b_2), (a_2, b_3)\}.$$

Este conjunto puede representarse gráficamente mediante un diagrama de árbol como el que se ilustra en la Figura 1.4 . •

**Ejemplo 1.7** Si un hombre tiene 6 camisas y 7 pantalones, ¿de cuántas maneras diferentes puede vestirse con estas prendas?

Respuesta:  $6 \times 7 = 42$ . •

En general los conjuntos producto  $A \times B$  y  $B \times A$  son distintos pues la pareja  $(a, b)$  es distinta de  $(b, a)$ , sin embargo ambos conjuntos tienen la misma cardinalidad, esto es, ambos tienen el mismo número de elementos. Si la cardinalidad de  $A$  es el número  $n$ , y la cardinalidad de  $B$  es  $m$ , entonces

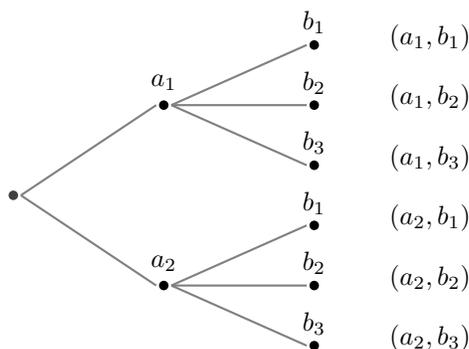


Figura 1.4: Diagrama de árbol.

la cardinalidad del conjunto  $A \times B$  es el producto  $n \cdot m$ . Este resultado es llamado principio de multiplicación y se aplica con mucha frecuencia en los procesos de conteo.

Un poco más generalmente, si  $A_1, A_2, \dots, A_k$  son conjuntos tales que  $\#A_i = n_i \geq 1$  para  $i = 1, \dots, k$ , entonces el producto Cartesiano  $A_1 \times A_2 \times \dots \times A_k$  que consta de todos los vectores de la forma  $(a_1, a_2, \dots, a_k)$  con  $a_i \in A_i$  tiene un total de  $n_1 \cdot n_2 \cdot \dots \cdot n_k$  elementos, es decir,

$$\#(A_1 \times \dots \times A_k) = n_1 \cdot n_2 \cdot \dots \cdot n_k.$$

**Ejemplo 1.8** Si una mujer tiene 3 sombreros, 6 blusas, 8 faldas y 10 pares de zapatos, ¿de cuántas formas diferentes puede vestirse usando una prenda de cada tipo? Respuesta:  $3 \times 6 \times 8 \times 10 = 1440$ . •

**Ejemplo 1.9** El producto Cartesiano  $\mathbb{R} \times \mathbb{R}$ , que es el conjunto de todas las parejas de números reales  $(x, y)$ , se le denota usualmente por  $\mathbb{R}^2$ . Análogamente se definen los conjuntos  $\mathbb{R}^3, \mathbb{R}^4, \dots, \mathbb{R}^n$ . •

Concluimos aquí nuestra rápida y breve revisión de la teoría elemental de conjuntos. Recordemos que estamos interesados en calcular probabilidades

de los diferentes eventos, es decir, de subconjuntos del espacio muestral que se obtienen de los diversos experimentos aleatorios. En la siguiente sección estudiaremos algunas formas de definir matemáticamente la probabilidad de un evento.

## Ejercicios

5. Sean  $A$ ,  $B$  y  $C$  tres subconjuntos arbitrarios de un espacio muestral  $\Omega$ . Exprese  $A \cup B \cup C$  como la unión de tres conjuntos ajenos. Utilice un diagrama de Venn para ilustrar su solución.
6. Use las propiedades básicas de las operaciones entre conjuntos para demostrar rigurosamente las siguientes igualdades. En cada caso dibuje un diagrama de Venn para ilustrar cada identidad.
  - a)  $A = (A \cap B) \cup (A \cap B^c)$ .
  - b)  $A^c - B^c = B - A$ .
  - c)  $A \cap B^c = A - (A \cap B)$ .
  - d)  $A \cup B = A \cup (B \cap A^c)$ .
  - e)  $(A - B) - C = A - (B - C)$ .
  - f)  $A - (B \cap C) = (A - B) \cup (A - C)$ .
7. Demuestre o proporcione un contraejemplo para las siguientes proposiciones. En cada caso dibuje un diagrama de Venn para ilustrar cada situación.
  - a)  $A \cap B \subseteq A \subseteq A \cup B$ .
  - b) Si  $A \cap B = \emptyset$  entonces  $A \subseteq B^c$ .
  - c) Si  $A \subseteq B$  entonces  $B^c \subseteq A^c$ .
  - d) Si  $A \cap B = \emptyset$  entonces  $A \cup B^c = B^c$ .
  - e) Si  $A \subseteq B$  entonces  $A \cup (B - A) = B$ .
  - f)  $(A^c \cap B) \cup (A \cap B^c) = (A \cap B)^c$ .

8. **Diferencia simétrica.** La diferencia simétrica entre los conjuntos  $A$  y  $B$  se puede también definir como sigue:

$$A\Delta B := (A - B) \cup (B - A).$$

Demuestre o proporcione un contraejemplo para las siguientes identidades:

- a)  $A\Delta B = (A \cup B) - (A \cap B)$ .
- b)  $A\Delta B = B\Delta A$ .
- c)  $A\Delta B = A^c \Delta B^c$ .
- d)  $(A\Delta B)\Delta C = A\Delta(B\Delta C)$ .
- e)  $A \cap (B\Delta C) = (A \cap B)\Delta(A \cap C)$ .
- f)  $A \cup (B\Delta C) = (A \cup B)\Delta(A \cup C)$ .

Sugerencia: Utilice un diagrama de Venn.

9. Sean  $A$ ,  $B$  y  $C$  tres eventos de un experimento aleatorio. Exprese las siguientes oraciones en términos de los conjuntos  $A$ ,  $B$  y  $C$ :

- a) Ocurre  $A$  o  $B$  pero no  $C$ .
- b) Ninguno de estos tres eventos ocurre.
- c) Sólo uno de ellos ocurre.
- d) Exactamente dos de ellos ocurren.
- e) Por lo menos uno de ellos ocurre.
- f) A lo sumo dos de ellos ocurren.

Sugerencia: Utilice un diagrama de Venn.

10. En una población humana en donde el número de mujeres duplica el número de hombres el 42% de los hombres son mayores de 50 años y el 38% de las mujeres son mayores de 50 años. ¿Qué porcentaje total de la población es mayor a 50 años?
11. **Función indicadora.** La función indicadora de un evento cualquiera  $A$  se denota por  $1_A : \Omega \rightarrow \mathbb{R}$  y toma el valor uno dentro del evento  $A$  y cero fuera de él. Demuestre que:

- a)  $1_{\Omega}(\omega) = 1$ .  
 b)  $1_{\emptyset}(\omega) = 0$ .  
 c)  $1_{A \cup B}(\omega) = 1_A(\omega) + 1_B(\omega) - 1_{A \cap B}(\omega)$ .  
 d)  $1_{\bigcap_{i=1}^n A_i}(\omega) = \prod_{i=1}^n 1_{A_i}(\omega)$ .  
 e) Si  $A_1, \dots, A_n$  son eventos ajenos dos a dos, entonces

$$1_{\bigcup_{i=1}^n A_i}(\omega) = \sum_{i=1}^n 1_{A_i}(\omega).$$

Sugerencia: Toda función indicadora separa su dominio en dos partes disjuntas. Verifique estas identidades en cada una de estas partes.

12. **Señales.** Se transmiten cuatro señales en un canal de comunicación. Debido al ruido aleatorio que se presenta en el canal, cada señal se recibe bien o con distorsión. Defina el evento  $D_i$  como aquél que indica que la  $i$ -ésima señal está distorsionada. Exprese los siguientes eventos en términos de los eventos  $D_i$ :
- a) Una señal está distorsionada.  
 b) Dos señales están distorsionadas.  
 c) A lo sumo hay dos señales consecutivas distorsionadas.  
 d) No hay dos señales consecutivas distorsionadas.  
 e) Por lo menos hay dos señales consecutivas distorsionadas.
13. Considere el experimento aleatorio de escoger al azar dos números  $x$  y  $y$  del intervalo unitario  $(0, 1)$ . El espacio muestral  $\Omega$  para este experimento es entonces el producto Cartesiano  $(0, 1) \times (0, 1)$ . Represente en un plano Cartesiano este espacio muestral e identifique los siguientes eventos:
- a)  $A = (x > 1/2) \cap (y < 1/2)$ .  
 b)  $B = (x < 2y) \cup (y < 1/2)$ .  
 c)  $C = (x^2 + y^2 < 1) - (y < x)$ .  
 d)  $D = (|x - y| < 1/4) \cup (|1 - x - y| < 1/4)$ .

14. Suponga que un experimento aleatorio consiste en escoger una persona al azar dentro de una población dada. Defina los eventos:

$H$  = “La persona escogida es hombre.”

$E$  = “La persona escogida cuenta con un empleo.”

$C$  = “La persona escogida es casada.”

Expresé en palabras el tipo de personas, según las características anteriores, determinadas por los siguientes eventos:

a)  $H \cap E$ .

e)  $H \cap E \cap C$ .

b)  $H^c \cap E^c$ .

f)  $(H \cap C) - E$ .

c)  $H \cup E$ .

g)  $(H - E) \cap C^c$ .

d)  $H - E$ .

h)  $C^c - E^c$ .

15. Un número entero es seleccionado al azar. Defina los eventos:

$A$  = “El número escogido es par.”

$B$  = “El número escogido termina en 5.”

$C$  = “El número escogido termina en 0.”

Describa con palabras los siguientes eventos:

a)  $A \cap C$ .

c)  $A \cap B$ .

b)  $B \cup C$ .

d)  $A - B$ .

16. **Circuitos.** Considere el diagrama de la Figura 1.5 el cual representa un circuito eléctrico. Los componentes  $A$ ,  $B_1$ ,  $B_2$ ,  $B_3$  y  $C$  pueden o no pueden funcionar. Denotemos por la misma letra el evento de que el correspondiente componente funcione y por su complemento el hecho de que no funcione. Sea  $F$  el evento de que el circuito completo funcione. Escriba  $F$  y  $F^c$  en términos de los eventos  $A$ ,  $B_1$ ,  $B_2$ ,  $B_3$  y  $C$ .

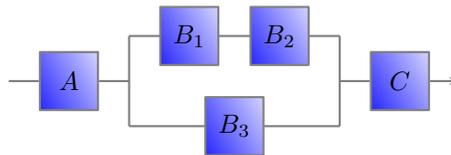


Figura 1.5

## 1.4. Probabilidad clásica

La probabilidad de un evento  $A$  es un número real en el intervalo  $[0, 1]$  que se denota por  $P(A)$  y representa una medida de la frecuencia con la que se observa la ocurrencia de dicho evento cuando se efectúa el experimento aleatorio en cuestión. Existen definiciones más específicas de la probabilidad, algunas de las cuales estudiaremos en las siguientes secciones.

**Definición 1.2** Sea  $A$  un subconjunto de un espacio muestral  $\Omega$  de cardinalidad finita. Se define la **probabilidad clásica** del evento  $A$  como el cociente:

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

en donde el símbolo  $\#A$  denota la cardinalidad o número de elementos del conjunto  $A$ .

Claramente esta definición es sólo válida para espacios muestrales finitos, pues forzosamente necesitamos suponer que el número de elementos en  $\Omega$  es finito. Además, el espacio  $\Omega$  debe ser equiprobable, pues para calcular la probabilidad de un evento  $A$ , únicamente necesitamos contar cuántos elementos tiene  $A$  respecto del total  $\Omega$ , sin importar exactamente cuáles elementos particulares sean. Por lo tanto, esta definición de probabilidad presupone que todos los elementos de  $\Omega$  son igualmente probables o tienen el mismo peso. Este es el caso por ejemplo del lanzamiento de un dado equilibrado. Para este experimento el espacio muestral es el conjunto  $\Omega =$

$\{1, 2, 3, 4, 5, 6\}$ , y si deseamos calcular la probabilidad (clásica) del evento  $A$  correspondiente a obtener un número par, es decir, la probabilidad de  $A = \{2, 4, 6\}$ , entonces

$$P(A) = \frac{\#\{2, 4, 6\}}{\#\{1, 2, 3, 4, 5, 6\}} = \frac{3}{6} = \frac{1}{2}.$$

Es inmediato verificar que esta forma de calcular probabilidades cumple, entre otras, las siguientes propiedades:

1.  $P(\Omega) = 1$ .
2.  $P(A) \geq 0$  para cualquier evento  $A$ .
3.  $P(A \cup B) = P(A) + P(B)$  cuando  $A$  y  $B$  son ajenos.

A esta forma de definir la probabilidad también se le conoce con el nombre de [probabilidad de Laplace](#) en honor del astrónomo y matemático francés Pierre-Simon Laplace, quien estableció de una manera sistemática y rigurosa los principios y propiedades de esta manera de calcular probabilidades. Más adelante retomaremos esta definición de probabilidad cuando revisemos algunas técnicas de conteo.

## Ejercicios

17. Sean  $A$  y  $B$  dos eventos cualesquiera de un experimento aleatorio con espacio muestral finito y equiprobable. Demuestre que la definición de la probabilidad clásica satisface las siguientes propiedades:
  - a)  $P(\emptyset) = 0$ .
  - b)  $P(A^c) = 1 - P(A)$ .
  - c) Si  $A \subseteq B$  entonces  $P(A) \leq P(B)$ .
  - d)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

18. El juego de una feria consiste en pedirle a un jugador que arroje al azar 4 monedas equilibradas, indistinguibles, todas ellas de una unidad monetaria y marcadas con “cara” y “cruz”. Cada moneda que caiga “cara” es recogida por el jugador y se le entrega una moneda adicional de la misma denominación como premio. Por otro lado, el jugador pierde cualquier moneda que caiga “cruz”.
- Determine todos los posibles montos de ganancias y pérdidas para este juego.
  - Calcule las probabilidades de todos los posible montos.
19. Un experimento aleatorio consiste en lanzar a un mismo tiempo dos dados equilibrados indistinguibles. Determine si a este experimento aleatorio se le puede asignar un espacio muestral finito y equiprobable.
20. **Puntos.** Suponga que un experimento aleatorio tiene como espacio muestral el conjunto de pares de números  $(x, y)$  tales que tanto  $x$  como  $y$  toman valores en el conjunto  $\{1, \dots, n\}$  y que se considera que cualquiera de estos puntos en el plano Cartesiano ocurre con idéntica probabilidad. Calcule la probabilidad de que al efectuar una vez el experimento aleatorio:
- se obtenga un punto en la diagonal, es decir,  $x = y$ .
  - se obtenga un punto en la orilla, es decir  $x = 1$  ó  $x = n$  ó  $y = 1$  ó  $y = n$ .
  - se obtenga un punto  $(x, y)$  tal que  $x \leq y$ .
  - se obtenga un punto  $(x, y)$  tal que  $|x - y| \leq 1$ .
21. Una moneda equilibrada y marcada con “cara” y “cruz” se lanza 4 veces consecutivas. Calcule la probabilidad de que el número de veces que cae “cara” sea estrictamente mayor al número de veces que cae “cruz”.

## 1.5. Probabilidad geométrica

Esta es una extensión de la definición de probabilidad clásica en donde ahora la probabilidad de un evento  $A$  se calcula ya no a través de su cardinalidad sino mediante la determinación de su área, volumen o alguna característica geométrica según el problema que se trate. Para el caso de áreas la definición es la siguiente.

**Definición 1.3** Si un experimento aleatorio tiene como espacio muestral  $\Omega \subset \mathbb{R}^2$  cuya área está bien definida y es finita, entonces se define la **probabilidad geométrica** de un evento  $A \subseteq \Omega$  como

$$P(A) = \frac{\text{Área de } A}{\text{Área de } \Omega},$$

cuando el concepto de área del subconjunto  $A$  está bien definido.

Para poder aplicar la fórmula anterior es necesario suponer que el espacio muestral es equiprobable en el sentido de que la probabilidad de observar la ocurrencia de un evento  $A$  depende únicamente de su área y no del conjunto mismo. Esta definición puede enunciarse también para el caso cuando  $\Omega$  es un subconjunto de  $\mathbb{R}$  y en tal caso no se habla de área sino de longitud. O bien cuando  $\Omega \subset \mathbb{R}^3$  y en ese caso se habla de volumen, etcétera. Ilustraremos la situación mediante algunos ejemplos.

**Ejemplo 1.10 (El problema del juego de una feria)** El juego de una feria consiste en lanzar monedas de radio  $r$  sobre un tablero cuadrado como el que se muestra en la Figura 1.6, en donde el lado de cada cuadrado mide  $a$  unidades. Un jugador se hace acreedor a un premio si la moneda lanzada no toca ninguna de las líneas. ¿De qué tamaño deben ser  $a$  y  $r$  para que la probabilidad de ganar en este juego sea menor a  $1/4$ ?

Solución: Primero debemos observar que es suficiente considerar lo que sucede únicamente en el cuadrado donde cae el centro de la moneda. No es difícil darse cuenta que la moneda no toca ninguna línea si su centro cae dentro del cuadrado interior que se muestra en la Figura 1.7.

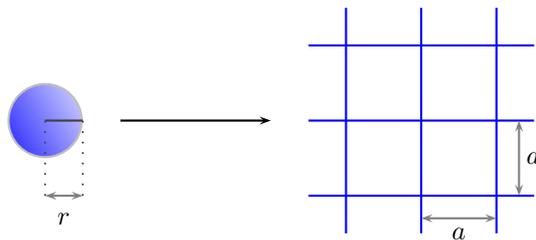


Figura 1.6

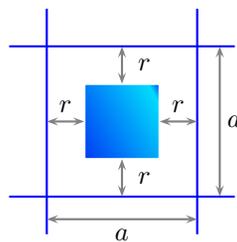


Figura 1.7

Por lo tanto, si  $A$  denota el evento de ganar con un lanzamiento en este juego, entonces la probabilidad de  $A$  es el cociente entre el área favorable y el área total, es decir,

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

Si deseamos que esta probabilidad sea menor a  $1/4$ , entonces de aquí puede uno encontrar que  $a$  y  $r$  deben cumplir la relación  $a < 4r$ . Cuando  $a = 4r$  la probabilidad de ganar es exactamente  $1/4$ . •

**Ejemplo 1.11 (El problema de los dos amigos)** Dos amigos deciden encontrarse en cierto lugar pero olvidan la hora exacta de la cita, únicamente recuerdan que la hora era entre las 12:00 y las 13:00 horas. Cada uno de ellos decide llegar al azar en ese lapso y esperar solamente 10 minutos. ¿Cuál es la probabilidad de que los amigos se encuentren?

Solución: Sean  $x$  y  $y$  el tiempo medido en minutos en los que llegan los dos amigos. El espacio muestral del experimento que consta de observar la hora de llegada de los dos amigos consiste de las parejas  $(x, y)$  tal que cada entrada de este vector es un instante entre las 12:00 y las 13:00 horas. Gráficamente el espacio muestral se puede representar como el cuadrado que se muestra en la Figura 1.8.

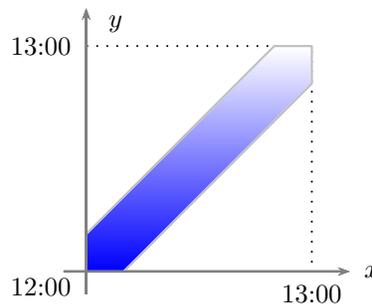


Figura 1.8

Los amigos se encuentran si  $x$  y  $y$  son tales que  $|x - y| \leq 10$  y esta región corresponde a la franja que se muestra en la Figura 1.8 y cuya área en minutos es  $(60)^2 - (50)^2$ . Si  $A$  denota el evento de interés, entonces tenemos que

$$P(A) = \frac{(60)^2 - (50)^2}{(60)^2} = \frac{36 - 25}{36} = \frac{11}{36}.$$

•

En la sección de ejercicios el lector encontrará algunos otros problemas de aplicación de la probabilidad geométrica. Para encontrar la solución a estos problemas y como sugerencia se recomienda al lector primeramente determinar con claridad el experimento aleatorio en cuestión y establecer con precisión un espacio muestral adecuado a la pregunta que se desea contestar. Observemos finalmente que la probabilidad geométrica también cumple, entre otras, con las siguientes propiedades:

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

2.  $P(A) \geq 0$  para cualquier evento  $A$ .
3.  $P(A \cup B) = P(A) + P(B)$  cuando  $A$  y  $B$  son ajenos.

## Ejercicios

22. Suponga que se tiene un experimento aleatorio con espacio muestral  $\Omega \subseteq \mathbb{R}^2$  cuya área está bien definida y es finita. Sean  $A$  y  $B$  dos eventos de este experimento cuya área está bien definida. Demuestre que la definición de la probabilidad geométrica satisface las siguientes propiedades:

- a)  $P(\emptyset) = 0$ .
- b)  $P(A^c) = 1 - P(A)$ .
- c) Si  $A \subseteq B$  entonces  $P(A) \leq P(B)$ .
- d)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

23. Se escogen dos números  $x$  y  $y$  al azar dentro del intervalo unitario  $[0, 1]$ . ¿Cuál es la probabilidad de que la suma de estos números sea mayor a uno y que al mismo tiempo la suma de sus cuadrados sea menor a uno?

Respuesta:  $(\pi - 2)/4$ .

24. Se escogen dos números al azar dentro del intervalo unitario  $[0, 1]$ . ¿Cuál es la probabilidad de que el producto de estos números sea menor a  $1/2$ ?

Respuesta:  $(\ln 2e)/2$ .

25. Se escogen dos números  $x$  y  $y$  al azar dentro del intervalo unitario  $[0, 1]$ . ¿Cuál es la probabilidad de que:

- a) la distancia de  $x$  a  $y$  sea menor a  $1/2$ ?
- b) la distancia de  $x$  a  $y$  sea mayor a  $1/4$ ?
- c) la distancia de  $x$  a cero sea a lo sumo la distancia de  $y$  a uno?

Respuestas: a)  $3/4$ . b)  $9/16$ . c)  $1/2$ .

26. Se escoge un número  $a$  al azar dentro del intervalo  $(-1, 1)$ . ¿Cuál es la probabilidad de que la ecuación cuadrática  $ax^2 + x + 1 = 0$  tenga dos raíces reales?

Respuesta:  $5/8$ .

27. Se escogen dos números  $b$  y  $c$  al azar dentro del intervalo unitario  $[0, 1]$ . ¿Cuál es la probabilidad de que la ecuación cuadrática  $x^2 + bx + c = 0$  tenga raíces complejas?

Respuesta:  $11/12$ .

28. **El problema de la aguja de Buffón.** Considere un conjunto infinito de líneas horizontales paralelas sobre una superficie plana como se muestra en la Figura 1.9. La distancia entre una línea y otra es  $L$ . Se deja caer una aguja de longitud  $\ell$  sobre la superficie. Suponga  $\ell \leq L$ . ¿Cuál es la probabilidad de que la aguja cruce alguna línea?

Respuesta:  $\frac{2\ell}{\pi L}$ .

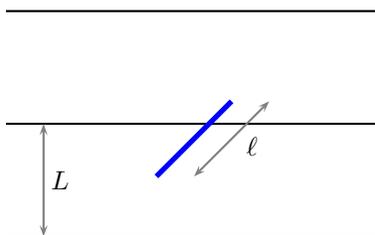


Figura 1.9

Como una simplificación suponga el caso  $\ell = L$  y sea  $A$  el evento que ocurre cuando la aguja toca alguna de las líneas. Si se efectúa este experimento  $n$  veces y  $n_A$  denota el número de ocurrencias del evento  $A$ , entonces el cociente  $n_A/n$  es cercano a  $P(A)$ . Así, tenemos la aproximación

$$\frac{2}{\pi} \approx \frac{n_A}{n},$$

de donde puede obtenerse una aproximación para  $\pi$ , a partir del experimento simple de lanzar agujas en una superficie de líneas paralelas:

$$\pi \approx \frac{2n}{n_A}.$$

Suponga ahora  $\ell \geq L$ . ¿Cuál es la probabilidad de que la aguja cruce alguna línea?

Respuesta:  $1 - \frac{2}{\pi} \arcsin\left(\frac{L}{\ell}\right) + \frac{2\ell}{\pi L} \left(1 - \cos\left(\arcsin\left(\frac{L}{\ell}\right)\right)\right)$ .

Usando la identidad  $\cos(\arcsin x) = \sqrt{1 - x^2}$ ,  $-1 \leq x \leq 1$ , la respuesta anterior se puede escribir también como sigue:

$$\frac{2\ell}{\pi L} - \frac{2}{\pi L} \left(\sqrt{\ell^2 - L^2} + L \arcsin\left(\frac{L}{\ell}\right)\right) + 1.$$

29. Se escogen al azar y de manera independiente tres números  $a$ ,  $b$  y  $c$  dentro del intervalo unitario  $[0, 1]$ . ¿Cuál es la probabilidad de que la suma de estos números sea menor a uno?

Respuesta:  $1/6$ .

30. [El problema de la varilla de metal](#). Una varilla de metal de longitud  $\ell$  se rompe en dos puntos distintos escogidos al azar. ¿Cuál es la probabilidad de que los tres segmentos así obtenidos formen un triángulo?

Respuesta:  $1/4$ .

31. Un pasajero llega en autobús a la estación de trenes. La hora de llegada del autobús es aleatoria entre las 9:00 y 10:00 hrs. Por otro lado, el tren que debe tomar el pasajero parte de la estación también al azar entre las 9:00 y 10:00 hrs. El pasajero podrá subirse al tren si el autobús llega por lo menos cinco minutos antes que el tren parta. ¿Cuál es la probabilidad de que el pasajero aborde el tren?

Respuesta:  $(11/12)^2/2$ .

32. Considere nuevamente el problema de la feria del Ejemplo 1.10 en la página 21. Suponga ahora que el jugador gana si la moneda toca a lo sumo un línea. ¿Cuál es la probabilidad de ganar?

Respuesta:  $1 - (2r/a)^2$ .

33. Dos personas tienen la misma probabilidad de llegar al lugar de su cita en cualquier instante dentro del intervalo de tiempo  $[0, T]$  y llegan de manera independiente una de otra. Encuentre la probabilidad de que el tiempo que una persona tenga que esperar a la otra sea a lo sumo  $t > 0$ .

$$\text{Respuesta: } \begin{cases} t(2T - t)/T^2 & \text{si } 0 \leq t < T, \\ 1 & \text{si } t \geq T. \end{cases}$$

34. Suponga que se escoge un punto al azar dentro de un segmento de recta de longitud  $L$  de tal forma que la probabilidad de que el punto caiga en un subsegmento es la misma de que caiga en cualquier otro subsegmento de la misma longitud. Calcule la probabilidad de que la distancia del punto al centro del segmento sea menor a  $\ell$ .

$$\text{Respuesta: } \begin{cases} \frac{2\ell}{L} & \text{si } 2\ell < L, \\ 1 & \text{si } 2\ell \geq L. \end{cases}$$

35. Suponga que se escogen al azar y de manera independiente dos puntos dentro de un segmento de recta de longitud  $L$  de tal forma que la probabilidad de que cualquiera de los puntos caiga en un subsegmento es la misma de que caiga en cualquier otro subsegmento de la misma longitud. Calcule la probabilidad de que por lo menos uno de los puntos caiga en la primera mitad del intervalo.

Respuesta:  $3/4$ .

36. Se escogen dos números al azar de manera independiente uno del otro dentro del intervalo  $[0, L]$ . Encuentre la probabilidad de que el promedio aritmético de estos dos números se encuentre dentro del subintervalo  $[a, b] \subseteq [0, L]$ .

$$\text{Respuesta: } \begin{cases} \frac{L^2 - 2(a^2 + (L-b)^2)}{L^2} & \text{si } a < L/2 < b, \\ \frac{2(b+a)(b-a)}{L^2} & \text{si } b \leq L/2, \\ \frac{2(2L-a-b)(b-a)}{L^2} & \text{si } a \geq L/2. \end{cases}$$

37. Se escogen al azar y de manera independiente tres números reales dentro del intervalo  $[0, L]$ . Calcule la probabilidad de que el promedio aritmético de estos números sea menor a  $L/3$ .

Respuesta:  $1/6$ .

38. **Triángulos 1.** Se escogen dos números  $x$  y  $y$  al azar de manera independiente uno del otro dentro del intervalo  $[0, \ell]$ . Calcule la probabilidad de que las longitudes  $x$ ,  $y$  y  $\ell$  formen un triángulo.

Respuesta:  $1/2$ .

39. **Triángulos 2.** Se escogen tres números  $x$ ,  $y$  y  $z$  al azar de manera independiente uno del otro dentro del intervalo  $[0, \ell]$ . Calcule la probabilidad de que las longitudes  $x$ ,  $y$  y  $z$  formen un triángulo.

*Sugerencia:* En un plano Cartesiano dibuje la curva de nivel  $z = z_0$ , por ejemplo para  $z_0 = \ell/2$ .

Respuesta:  $1/2$ .

40. Se escoge un punto  $(x, y, z)$  al azar dentro del cubo  $[-1, 1] \times [-1, 1] \times [-1, 1]$ , de manera uniforme, es decir, con la misma probabilidad para dos regiones con el mismo volumen. Calcule la probabilidad del evento

$$A = \{(x, y, z) \in \Omega : |x| + |y| + |z| \leq 1\}.$$

Respuesta:  $1/6$ .

## 1.6. Probabilidad frecuentista

Suponga que se realizan  $n$  repeticiones de un cierto experimento aleatorio y se registra el número de veces que ocurre un determinado evento  $A$ . Esta información puede ser usada de la siguiente forma para definir la probabilidad de  $A$ .

**Definición 1.4** Sea  $n_A$  el número de ocurrencias de un evento  $A$  en  $n$  realizaciones de un experimento aleatorio. La **probabilidad frecuentista** del evento  $A$  se define como el límite

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

En este caso, debemos hacer notar que no es humanamente posible llevar a cabo una infinidad de veces el experimento aleatorio, de modo que en la práctica no es posible encontrar mediante este mecanismo y de manera exacta la probabilidad de un evento cualquiera, aunque permite tener una aproximación del valor de  $P(A)$ . Esta limitación hace que esta definición de probabilidad no sea enteramente formal, pero tiene algunas ventajas. Veamos un ejemplo concreto. Consideremos nuevamente el experimento aleatorio de lanzar un dado equilibrado y registrar la ocurrencia del evento  $A$

definido como el conjunto  $\{2, 4, 6\}$ . Después de lanzar el dado 20 veces se obtuvieron los siguientes resultados:

Núm.	Resultado	$n_A/n$	Núm.	Resultado	$n_A/n$
1	3	0/1	11	2	7/11
2	6	1/2	12	5	7/12
3	2	2/3	13	1	7/13
4	1	2/4	14	6	8/14
5	4	3/5	15	3	8/15
6	6	4/6	16	1	8/16
7	3	4/7	17	5	8/17
8	4	5/8	18	5	8/18
9	2	6/9	19	2	9/19
10	5	6/10	20	6	10/20

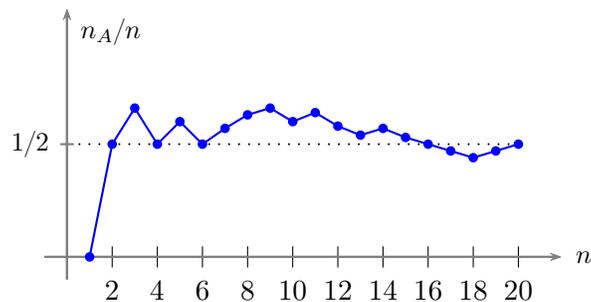


Figura 1.10

En la gráfica de la Figura 1.10 se muestra el singular comportamiento de este cociente a lo largo del tiempo, al principio se pueden presentar algunas oscilaciones pero eventualmente el cociente parece estabilizarse en un cierto valor. Realizando un mayor número de observaciones del experimento, no es difícil creer que el cociente  $n_A/n$  se estabiliza en  $1/2$  cuando el dado está equilibrado y el número de ensayos  $n$  es grande. Se invita al lector intrigado a efectuar un experimento similar y corroborar esta interesante [regularidad estadística](#) con éste o cualquier otro experimento aleatorio de su interés. Más adelante formalizaremos este resultado mediante la así llamada

ley de los grandes números.

**Simulación 1.1** Arroje cien veces una moneda y registre los resultados en una lista. Calcule y grafique los cocientes  $n_A/n$  cuando el evento  $A$  corresponde a obtener alguna de las caras de la moneda. ¿Converge el cociente  $n_A/n$  a  $1/2$ ? Para agilizar el experimento puede usted dividir los cien lanzamientos en varios grupos de personas y después juntar los resultados. Alternativamente, investigue la forma de simular este experimento en una computadora y calcule  $n_A/n$  para distintos valores de  $n$ . Véase la sección de ejercicios. •

Es fácil comprobar que la probabilidad frecuentista también cumple las siguientes propiedades que ya hemos mencionado antes:

1.  $P(\Omega) = 1$ .
2.  $P(A) \geq 0$  para cualquier evento  $A$ .
3.  $P(A \cup B) = P(A) + P(B)$  cuando  $A$  y  $B$  son ajenos.

## Ejercicios

41. Sean  $A$  y  $B$  dos eventos de un experimento aleatorio. Demuestre que la definición de la probabilidad frecuentista satisface las siguientes propiedades:
  - a)  $P(\emptyset) = 0$ .
  - b)  $P(A^c) = 1 - P(A)$ .
  - c) Si  $A \subseteq B$  entonces  $P(A) \leq P(B)$ .
  - d)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
42. **Moneda.** El lanzamiento de una moneda puede simularse en R usando el comando que aparece abajo. Realice 100 simulaciones del lanzamiento de una moneda equilibrada y compruebe experimentalmente

que el número de veces que aparece una de las caras entre el total de lanzamientos se aproxima a  $1/2$  conforme el número de lanzamientos crece.

```
# Simulación de 20 lanzamientos de una moneda equilibrada  
> sample(0:1, 20, replace=TRUE)  
[1] 0 1 0 1 0 0 1 1 1 0 1 0 0 0 0 1 1 1 0 1
```

43. **Dado.** El lanzamiento de un dado puede simularse en R usando el comando que aparece abajo. Realice 100 simulaciones del lanzamiento de un dado equilibrado y compruebe experimentalmente que el número de veces que aparece una de las caras entre el total de lanzamientos se aproxima a  $1/6$  conforme el número de lanzamientos crece.

```
# Simulación de 20 lanzamientos de un dado equilibrado  
> sample(1:6, 20, replace=TRUE)  
[1] 1 5 3 4 4 6 3 4 3 3 2 4 4 5 3 1 6 5 6 6
```

## 1.7. Probabilidad subjetiva

En este caso la probabilidad de un evento depende del observador, es decir, depende de lo que el observador conoce del fenómeno en estudio. Puede parecer un tanto informal y poco seria esta definición de la probabilidad de un evento, sin embargo en muchas situaciones es necesario recurrir a un experto para tener por lo menos una idea vaga de cómo se comporta el fenómeno de nuestro interés y saber si la probabilidad de un evento es alta o baja. Por ejemplo, ¿cuál es la probabilidad de que un cierto equipo de fútbol gane en su próximo partido? Ciertas circunstancias internas del equipo, las condiciones del equipo rival o cualquier otra condición externa, son elementos que sólo algunas personas conocen y que podrían darnos una idea más exacta de esta probabilidad. Esta forma subjetiva de asignar probabilidades a los distintos eventos debe, sin embargo, ser consistente con una serie de reglas naturales que estudiaremos a continuación.

## 1.8. Probabilidad axiomática

En la definición axiomática de la probabilidad no se establece la forma explícita de calcular las probabilidades sino únicamente se proponen las reglas que el cálculo de probabilidades debe satisfacer. Los siguientes tres postulados o axiomas fueron establecidos en 1933 por el matemático ruso Andrey Nikolaevich Kolmogorov.

### Axiomas de la probabilidad

1.  $P(A) \geq 0$ .
2.  $P(\Omega) = 1$ .
3.  $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$  cuando  $A_1, A_2, \dots$  son ajenos dos a dos.

Recordemos que un [axioma](#) o [postulado](#) es una proposición que se acepta como válida y sobre la cual se funda una teoría, en este caso la teoría de la probabilidad. En particular, estos postulados han sido tomados directamente del análisis cuidadoso y reflexivo de las definiciones de probabilidad mencionadas anteriormente. Y no es difícil verificar que las definiciones anteriores de probabilidad satisfacen estos tres axiomas. Por ejemplo, considerando nuevamente la definición de probabilidad frecuentista, se debe realizar una sucesión de  $n$  ensayos de un experimento aleatorio y calcular el cociente  $n_A/n$  para un evento  $A$  cualquiera. Se observa claramente que el cociente  $n_A/n$  es no negativo, esto es el primer axioma. Si  $A$  es el evento total  $\Omega$ , entonces  $n_\Omega = n$  y por lo tanto  $n_\Omega/n = 1$ , esto es el segundo axioma. Finalmente, si  $A$  y  $B$  son dos eventos ajenos, se tiene que  $n_{A \cup B} = n_A + n_B$  y por lo tanto

$$\frac{n_{A \cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n},$$

de donde surge el tercer axioma para una colección numerable de eventos ajenos dos a dos. Tenemos a continuación la definición axiomática de la probabilidad.

**Definición 1.5** A cualquier función  $P$  definida sobre una colección de eventos que satisfaga los tres axiomas de Kolmogorov se le llama **medida de probabilidad**, o simplemente **probabilidad**.

En esta definición no hemos sido muy precisos en definir el dominio de la función  $P$ , sólo hemos dicho que tal dominio es una colección de eventos. En la siguiente sección especificaremos con mas detalle a esta colección de eventos, por ahora nos interesa estudiar las propiedades generales que la función  $P$  pueda tener. Tomando como base los axiomas de Kolmogorov y usando la teoría elemental de conjuntos demostraremos que toda medida de probabilidad cumple con una serie de propiedades interesantes.

**Proposición 1.1**  $P(\emptyset) = 0$ .

**Demostración.** Tomando  $A_1 = A_2 = \dots = \emptyset$  en el tercer axioma tenemos que

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots .$$

La única solución de esta ecuación es  $P(\emptyset) = 0$ . •

**Proposición 1.2** Sea  $A_1, A_2, \dots, A_n$  una colección finita de eventos ajenos dos a dos. Entonces

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

**Demostración.** Definiendo  $A_{n+1} = A_{n+2} = \dots = \emptyset$  se verifica que esta sucesión infinita sigue siendo ajena dos a dos y por lo tanto, usando el tercer

axioma, tenemos que

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

•

En particular, cuando  $A \cap B = \emptyset$  se cumple la identidad

$$P(A \cup B) = P(A) + P(B).$$

**Proposición 1.3** Para cualquier evento  $A$ ,  $P(A^c) = 1 - P(A)$ .

**Demostración.** De la teoría elemental de conjuntos tenemos que  $\Omega = A \cup A^c$ , en donde  $A$  y  $A^c$  son eventos ajenos. Aplicando el tercer axioma tenemos que

$$\begin{aligned} 1 &= P(\Omega) \\ &= P(A \cup A^c) \\ &= P(A) + P(A^c). \end{aligned}$$

•

La proposición recién demostrada establece que los eventos  $A$  y  $A^c$  tienen probabilidad complementaria, es decir, la suma de las probabilidades de estos eventos es siempre uno. Esta sencilla propiedad es bastante útil pues en ocasiones es más fácil calcular la probabilidad del complemento de un evento que del evento mismo. Más adelante tendremos múltiples ocasiones para aplicar este resultado.

Las siguientes dos proposiciones suponen la situación  $A \subseteq B$ , la cual se muestra gráficamente en la Figura 1.11. En particular, el siguiente resultado establece que la probabilidad es una función monótona no decreciente.

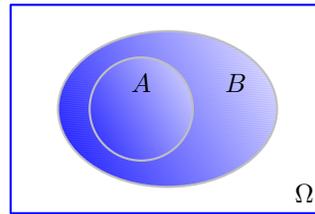


Figura 1.11:  $A \subseteq B$ .

**Proposición 1.4** Si  $A \subseteq B$ , entonces  $P(A) \leq P(B)$ .

**Demostración.** Primeramente escribimos  $B = A \cup (B - A)$ . Como  $A$  y  $B - A$  son eventos ajenos, por el tercer axioma,  $P(B) = P(A) + P(B - A)$ . Usando el primer axioma concluimos que  $P(B) - P(A) = P(B - A) \geq 0$ . De aquí obtenemos  $P(B) - P(A) \geq 0$ . •

**Proposición 1.5** Si  $A \subseteq B$ , entonces  $P(B - A) = P(B) - P(A)$ .

**Demostración.** Como  $B = A \cup (B - A)$ , siendo esta unión ajena, por el tercer axioma tenemos que  $P(B) = P(A) + P(B - A)$ . •

Por ejemplo, suponga que  $A$  y  $B$  son eventos tales que  $A \subseteq B$ ,  $P(A^c) = 0.9$  y  $P(B^c) = 0.6$ . Deseamos calcular  $P(B - A)$ . En esta situación es válida la fórmula  $P(B - A) = P(B) - P(A)$ , en donde,  $P(A) = 0.1$  y  $P(B) = 0.4$ . Por lo tanto,  $P(B - A) = 0.4 - 0.1 = 0.3$ . Observe el lector que en este

ejemplo sencillo no se especifica el experimento aleatorio en cuestión ni tampoco se definen los eventos  $A$  y  $B$ . El tratamiento es completamente analítico y los resultados son válidos para cualesquiera eventos  $A$  y  $B$  con las características señaladas.

**Proposición 1.6** Para cualquier evento  $A$ ,  $0 \leq P(A) \leq 1$ .

*Demostración.* Como  $A \subseteq \Omega$ , se tiene que  $P(A) \leq P(\Omega) = 1$ . La primera desigualdad,  $0 \leq P(A)$ , es simplemente el primer axioma. •

En palabras, la proposición anterior establece que la medida de probabilidad es una función que toma valores únicamente en el intervalo  $[0, 1]$ . El siguiente resultado proporciona una fórmula general para la probabilidad de la unión de cualesquiera dos eventos.

**Proposición 1.7** Para cualesquiera eventos  $A$  y  $B$ ,

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

*Demostración.* Primeramente observamos que para cualesquiera eventos  $A$  y  $B$  se cumple la igualdad  $A - B = A - (A \cap B)$ , en donde  $A \cap B \subseteq A$ , de modo que  $P(A - (A \cap B)) = P(A) - P(A \cap B)$ . Entonces escribimos a  $A \cup B$  como la unión disjunta de los siguientes tres eventos:

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

Ahora aplicamos la probabilidad. Por el tercer axioma,

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

•

En la Figura 1.12(a) el lector puede comprobar la validez de la fórmula anterior identificando las tres regiones ajenas de las que consta  $A \cup B$ . El término  $P(A)$  abarca las primeras dos regiones de izquierda a derecha,  $P(B)$  abarca la segunda y tercera región. Observe entonces que la región central ha sido contada dos veces de modo que el término  $-P(A \cap B)$  da cuenta de ello. De esta forma las tres regiones son contadas una sola vez y el resultado es la probabilidad del evento  $A \cup B$ .

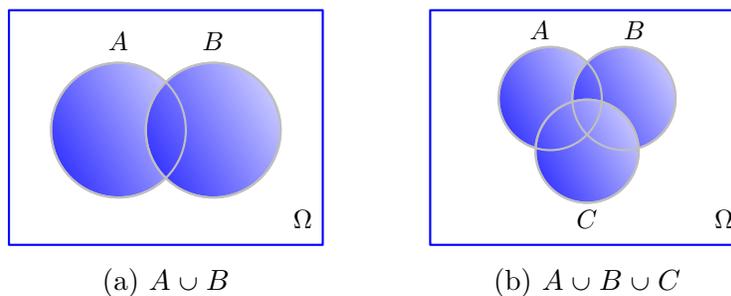


Figura 1.12

**Ejemplo 1.12** Sean  $A$  y  $B$  eventos ajenos tales que  $P(B) = 0.3$  y  $P(A \cap B^c) = 0.2$ . Encuentre  $P(A \cup B)$ .

Solución: Usaremos la fórmula  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Conocemos  $P(B)$ . Además  $P(A \cap B)$  es cero pues por hipótesis los eventos son ajenos, y  $P(A) = P(A \cap B^c) = 0.2$ . ¿Por qué? Por lo tanto,  $P(A \cup B) = 0.2 + 0.3 = 0.5$ . •

Como hemos mencionado, la fórmula anterior para la probabilidad de la unión de dos eventos es válida para cualesquiera que sean estos eventos, sin embargo, cuando los eventos son ajenos, es decir, cuando  $A \cap B = \emptyset$ , entonces la fórmula demostrada se reduce al tercer axioma de la probabilidad, es decir,  $P(A \cup B) = P(A) + P(B)$ . El siguiente resultado es una generalización del anterior e involucra tres eventos arbitrarios. La fórmula que a continuación se demuestra puede también verificarse usando el diagrama

de Venn que aparece en la Fig 1.12(b). Para ello siga los términos del lado derecho de la fórmula y compruebe que cada región es contada una sola vez de modo que el resultado final es la probabilidad del evento  $A \cup B \cup C$ .

**Proposición 1.8** Para cualesquiera eventos  $A$ ,  $B$  y  $C$ ,

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

**Demostración.** Agrupando adecuadamente y usando la fórmula para dos eventos,

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

•

Las propiedades anteriores son parte del estudio teórico y general de la probabilidad. En general, supondremos que la forma explícita de calcular estos números es conocida, o que se puede suponer un cierto modelo para llevar a cabo estos cálculos dependiendo del experimento aleatorio en cuestión. Por ejemplo, cuando el espacio muestral es finito y cada resultado puede suponerse igualmente probable, entonces usaremos la definición clásica de probabilidad. En otras situaciones asignaremos probabilidades de acuerdo a ciertos modelos conocidos. Esperamos que, a partir de las propiedades enunciadas y demostradas, el lector haya desarrollado cierta habilidad e intuición para escribir la demostración de alguna otra propiedad de la probabilidad. Se debe también mencionar que las demostraciones no son únicas y que es muy probable que el lector pueda producir alguna demostración

diferente a las que aquí se han presentado. Algunas otras propiedades de la probabilidad pueden encontrarse en la sección de ejercicios.

## Ejercicios

44. **Otras propiedades de la medida de probabilidad.** Demuestre las siguientes afirmaciones:

a)  $P(A \cap B \cap C) \leq P(A \cap B) \leq P(A)$ .

b)  $P(A) \leq P(A \cup B) \leq P(A \cup B \cup C)$ .

c)  $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$ .

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

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

f)  $P(A \Delta B) = P(A) + P(B) - 2P(A \cap B)$ .

g) Si  $A = B$  entonces  $P(A \Delta B) = 0$ .

h)  $P(A_1 \cap A_2) \geq 1 - P(A_1^c) - P(A_2^c)$ .

i)  $P(A_1 \cap A_2 \cap \dots \cap A_n) \geq 1 - \sum_{i=1}^n P(A_i^c)$ .

j) Si  $A_1 \cap A_2 \subseteq A$  entonces  $P(A) \geq P(A_1) + P(A_2) - 1$ .

k) Si  $\bigcap_{i=1}^n A_i \subseteq A$  entonces  $P(A) \geq \sum_{i=1}^n P(A_i) - (n - 1)$ .

l) Si  $A_1 \cap A_2 \subseteq A$  entonces  $P(A^c) \leq P(A_1^c) + P(A_2^c)$ .

m) Si  $\bigcap_{i=1}^n A_i \subseteq A$  entonces  $P(A^c) \leq \sum_{i=1}^n P(A_i^c)$ .

n) Si  $A_1, \dots, A_n$  tienen probabilidad cero entonces  $P\left(\bigcup_{i=1}^n A_i\right) = 0$ .

$\tilde{n}$ ) Si  $A_1, \dots, A_n$  tienen probabilidad uno entonces  $P\left(\bigcap_{i=1}^n A_i\right) = 1$ .

o)  $\max\{P(A_1), P(A_2)\} \leq P(A_1 \cup A_2) \leq \min\{P(A_1) + P(A_2), 1\}$ .

$$p) \max\{P(A_1), \dots, P(A_n)\} \leq P\left(\bigcup_{i=1}^n A_i\right) \leq \min\left\{\sum_{i=1}^n P(A_i), 1\right\}.$$

45. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- a) Si  $P(A) = 0$  entonces  $A = \emptyset$ .
- b) Si  $P(A) = 1$  entonces  $A = \Omega$ .
- c) Si  $P(A) = 0$  entonces  $P(A \cap B) = 0$ .
- d) Si  $P(A) = P(B)$  entonces  $A = B$ .
- e) Si  $P(A) \leq P(B)$  entonces  $A \subseteq B$ .
- f) Si  $P(A) = P(B) = p$  entonces  $P(A \cap B) \leq p^2$ .
- g) Si  $A \cap B = \emptyset$  entonces  $P(A) \leq P(B^c)$ .
- h) Si  $P(A) = P(B) = 0$  entonces  $P(A \cup B) = 0$ .
- i) Si  $P(A) = P(B) = 1$  entonces  $P(A \cap B) = 1$ .
- j) Si  $A \cap B \subseteq C$  entonces  $P(C^c) \leq P(A^c) + P(B^c)$ .
- k) Si  $P(A) > 1/2$  y  $P(B) > 1/2$  entonces  $P(A \cap B) > 0$ .

46. Suponga que un cierto experimento aleatorio tiene como espacio muestral el conjunto finito  $\Omega = \{1, \dots, n\}$ . Determine si las siguientes funciones son medidas de probabilidad: para cualquier  $A \subseteq \Omega$  se define

- a)  $P(A) = \sum_{k \in A} \frac{2k}{n(n+1)}$ .
- b)  $P(A) = \sum_{k \in A} \frac{2^k}{2^n - 2}$ ,  $n > 1$ .
- c)  $P(A) = \prod_{k \in A} \frac{k(n+1)}{k+1}$ .

47. Sean  $A$  y  $B$  eventos tales que  $P(A) = p$ ,  $P(B) = q$  y  $P(A \cap B) = r$ . Encuentre:

- a)  $P(A \cap B^c)$ .  
 b)  $P(A^c \cap B)$ .
- c)  $P(A^c \cap B^c)$ .  
 d)  $P(A \Delta B)$ .

48. Sea  $A$ ,  $B$  y  $C$  tres eventos tales que  $P(A) = P(B) = 1/3$ ,  $P(C) = 1/4$ ,  $P(A \cap B) = 1/6$  y  $P(B \cap C) = 0$ . Encuentre  $P(A \cap B \cap C)$ .

Respuesta:  $15/24$ .

49. **Fórmula de inclusión-exclusión** Use el método de inducción para demostrar que si  $A_1, A_2, \dots, A_n$  son eventos arbitrarios, entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{\substack{i,j,k \\ \text{distintos}}} P(A_i \cap A_j \cap A_k) - \dots + (-1)^n P(A_1 \cap \dots \cap A_n).$$

50. Demuestre que la probabilidad de que exactamente uno de los eventos  $A$  o  $B$  ocurra es  $P(A) + P(B) - 2P(A \cap B)$ .

51. Sean  $A$ ,  $B$  y  $C$  tres eventos. Demuestre que la probabilidad de que exactamente:

- a) uno de estos eventos ocurra es  
 $P(A) + P(B) + P(C) - 2P(A \cap B) - 2P(A \cap C) - 2P(B \cap C) + 3P(A \cap B \cap C)$ .
- b) dos de estos eventos ocurran es  
 $P(A \cap B) + P(A \cap C) + P(B \cap C) - 3P(A \cap B \cap C)$ .
- c) tres de estos eventos ocurran es  $P(A \cap B \cap C)$ .

¿Cuál es el evento que se obtiene al unir estas tres condiciones?

52. Sean  $P$  y  $Q$  dos medidas de probabilidad definidas sobre la misma colección de eventos. Demuestre que para cada  $\alpha \in [0, 1]$  la función  $\alpha P + (1 - \alpha)Q$  es también una medida de probabilidad.

## 1.9. Sigmas álgebras

En esta breve sección vamos a definir un conjunto que agrupe a todos los eventos de un mismo experimento aleatorio para los cuales se pueda definir o calcular sus probabilidades. En este texto elemental no haremos énfasis en este tipo de colecciones de eventos pero son fundamentales para un tratamiento serio de la teoría matemática de la probabilidad.

**Definición 1.6** Una colección  $\mathcal{F}$  de subconjuntos de un espacio muestral  $\Omega$  es una **álgebra** si cumple las siguientes tres condiciones:

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

Veamos con más detenimiento las condiciones que aparecen en la definición anterior. La primera condición establece que el espacio muestral en su totalidad debe pertenecer a la colección  $\mathcal{F}$ . Ciertamente el espacio muestral es un evento que siempre ocurre y como hemos visto su probabilidad está bien definida y es uno. La segunda condición asegura que si algún subconjunto  $A$  es de interés y por lo tanto se le considera un evento, entonces el complemento de tal conjunto también debe ser un evento. Nuevamente, por lo estudiado antes, la probabilidad del evento  $A^c$  está siempre dada por  $1 - P(A)$ . Finalmente el tercer requisito en la definición establece que si se tiene una sucesión finita de eventos entonces la unión de todos ellos también debe ser un evento, el cual corresponde a la ocurrencia de por lo menos uno de los eventos de la sucesión. Cuando esta tercera condición es también válida para sucesiones infinitas de eventos, a esta colección de subconjuntos de  $\Omega$  se le llama  $\sigma$ -álgebra, en donde el prefijo  $\sigma$  se refiere a la operación infinita involucrada.

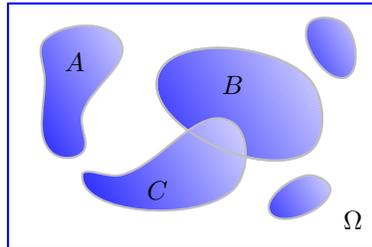


Figura 1.13

**Definición 1.7** Una colección  $\mathcal{F}$  de subconjuntos de un espacio muestral  $\Omega$  es una  $\sigma$ -álgebra si cumple las siguientes tres condiciones:

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

A los elementos de  $\mathcal{F}$  se les llama **eventos**.

Por lo tanto, a partir de ahora no llamaremos **evento** a cualquier subconjunto del espacio muestral sino únicamente a aquellos elementos que pertenezcan a una  $\sigma$ -álgebra asociada al espacio muestral. En el siguiente ejemplo se ilustra el hecho de que pueden existir varias  $\sigma$ -álgebras asociadas a un mismo espacio muestral.

**Ejemplo 1.13** Sea  $\Omega$  el espacio muestral de un experimento aleatorio. No es difícil comprobar que las siguientes colecciones de subconjuntos de  $\Omega$  son  $\sigma$ -álgebras:

1.  $\mathcal{F} = \{\emptyset, \Omega\}$ .
2.  $\mathcal{F} = \{A, A^c, \Omega, \emptyset\}$ , en donde  $A \subseteq \Omega$ .
3.  $\mathcal{F} = 2^\Omega$ .

•

De esta forma en la  $\sigma$ -álgebra  $\mathcal{F}$  se agrupa a todos los subconjuntos de  $\Omega$  para los que estamos interesados en calcular su probabilidad y tal colección constituye el dominio sobre el cual se define una medida de probabilidad. Así, a cada experimento aleatorio particular se le puede asociar una pareja  $(\Omega, \mathcal{F})$  compuesta por el espacio muestral y una  $\sigma$ -álgebra de eventos.

## Ejercicios

53. Sean  $A$  y  $B$  dos eventos. Demuestre que los siguientes conjuntos también son eventos, es decir, pertenecen a la misma  $\sigma$ -álgebra.

- |                  |                   |                       |
|------------------|-------------------|-----------------------|
| a) $\emptyset$ . | c) $A - B$ .      | e) $A \Delta B$ .     |
| b) $A \cap B$ .  | d) $A \cup B^c$ . | f) $A - (A \cap B)$ . |

54. Sean  $A$  y  $B$  dos subconjuntos arbitrarios del espacio muestral  $\Omega$  de un experimento aleatorio. ¿Qué conjuntos es necesario añadir a la colección  $\{A, B\}$  para convertirla en la mínima  $\sigma$ -álgebra de subconjuntos de  $\Omega$  que contiene a  $A$  y a  $B$ ?

55. Demuestre que toda  $\sigma$ -álgebra es álgebra. Mediante un contraejemplo demuestre que el recíproco es falso. Véase sin embargo el siguiente ejercicio.

56. Sea  $\Omega$  un espacio muestral finito. Demuestre que toda álgebra de subconjuntos de  $\omega$  es también una  $\sigma$ -álgebra.

57. Sean  $\mathcal{F}_1$  y  $\mathcal{F}_2$  dos  $\sigma$ -álgebras de subconjuntos de  $\Omega$ .

- a) Escriba la definición de  $\mathcal{F}_1 \cap \mathcal{F}_2$ .
- b) Demuestre que  $\mathcal{F}_1 \cap \mathcal{F}_2$  es una  $\sigma$ -álgebra.
- c) Mediante un contraejemplo demuestre que  $\mathcal{F}_1 \cup \mathcal{F}_2$  no necesariamente es una  $\sigma$ -álgebra.

## 1.10. Sigma álgebra de Borel

En esta sección estudiaremos brevemente un ejemplo de  $\sigma$ -álgebra que nos será de utilidad. Tomaremos como espacio muestral el conjunto de números reales  $\mathbb{R}$  y consideraremos la colección de intervalos de la forma  $(-\infty, x]$  para cualquier número real  $x$ , es decir, sea

$$\mathcal{C} = \{(-\infty, x] : x \in \mathbb{R}\}.$$

Esta colección contiene un número infinito no numerable de elementos pero es claro que no constituye una  $\sigma$ -álgebra de subconjuntos de  $\mathbb{R}$  pues, por ejemplo, no es cerrada bajo la operación de tomar complementos. La  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  se puede construir a partir de esta colección considerando la  $\sigma$ -álgebra más pequeña de subconjuntos de  $\mathbb{R}$  que contiene a la colección  $\mathcal{C}$ .

**Definición 1.8** La  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  se denota por  $\mathcal{B}(\mathbb{R})$  y se define como la mínima  $\sigma$ -álgebra de subconjuntos de  $\mathbb{R}$  que contiene a todos los intervalos de la forma  $(-\infty, x]$ , esto se escribe de la forma siguiente:

$$\mathcal{B}(\mathbb{R}) := \sigma\{(-\infty, x] : x \in \mathbb{R}\}.$$

A los elementos de  $\mathcal{B}(\mathbb{R})$  se les llama **conjuntos de Borel**, **conjuntos Borel medibles** o simplemente **Borelianos** de  $\mathbb{R}$ .

El símbolo  $\sigma$  en la expresión anterior significa que se está tomando la mínima  $\sigma$ -álgebra generada por la colección  $\{(-\infty, x] : x \in \mathbb{R}\}$ , y el adjetivo mínimo significa que si  $\mathcal{F}$  es una  $\sigma$ -álgebra que contiene a la colección  $\mathcal{C}$ , entonces  $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{F}$ , es decir,  $\mathcal{B}(\mathbb{R})$  es la más pequeña. Existen otras formas equivalentes de definir a  $\mathcal{B}(\mathbb{R})$  pero la que hemos presentado es suficiente para los propósitos de este texto. Es interesante mencionar también que el concepto de  $\sigma$ -álgebra de Borel puede extenderse a  $\mathbb{R}^n$  y aun a espacios más generales.

Para entender mejor a la  $\sigma$ -álgebra de Borel de  $\mathbb{R}$ , mostraremos a continuación algunos otros elementos que pertenecen a esta colección de subconjuntos de  $\mathbb{R}$ . Expresaremos a estos elementos como el resultado de operaciones

de elementos que sabemos que pertenecen a  $\mathcal{B}(\mathbb{R})$  y en donde las operaciones son aquellas bajo las cuales toda  $\sigma$ -álgebra es cerrada según la definición. En la siguiente tabla,  $x$  y  $y$  son cualesquiera números reales tales que  $x \leq y$ .

Subconjunto	Se expresa como
$(x, \infty)$	$(-\infty, x]^c$
$(-\infty, x)$	$\bigcup_{n=1}^{\infty} (-\infty, x - 1/n]$
$[x, \infty)$	$(-\infty, x)^c$
$[x, y]$	$(-\infty, y] - (-\infty, x)$
$[x, y)$	$(-\infty, y) - (-\infty, x)$
$(x, y]$	$(-\infty, y] - (-\infty, x]$
$(x, y)$	$(-\infty, y) - (-\infty, x]$
$\{x\}$	$(-\infty, x] - (-\infty, x)$
$\mathbb{N}$	$\bigcup_{n=1}^{\infty} \{n\}$
$\mathbb{Z}$	$\bigcup_{n=-\infty}^{\infty} \{n\}$
$\mathbb{Q}$	$\bigcup_{n,m=-\infty}^{\infty} \{n/m\}$
$\mathbb{I}$ (Irracionales)	$\mathbb{Q}^c$

Así, todos los conjuntos en la lista anterior son elementos de  $\mathcal{B}(\mathbb{R})$ , añadidos con los que se puedan producirse a partir de ellos. Uno podría entonces pensar que todo subconjunto de  $\mathbb{R}$  es un Boreliano pero ello no es cierto, se pueden construir subconjuntos de  $\mathbb{R}$  que no son Borelianos. Una vez que estudiemos en el siguiente capítulo el concepto de variable aleatoria, estaremos considerando siempre implícitamente conjuntos de Borel de  $\mathbb{R}$ .

### 1.11. Espacios de probabilidad

Con los elementos antes estudiados podemos ahora definir formalmente la estructura matemática diseñada para modelar un experimento aleatorio.

**Definición 1.9** Un espacio de probabilidad es una terna  $(\Omega, \mathcal{F}, P)$  en donde  $\Omega$  es un conjunto arbitrario,  $\mathcal{F}$  es una  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y  $P$  es una medida de probabilidad definida sobre  $\mathcal{F}$ .

El conjunto arbitrario  $\Omega$  representa usualmente el espacio muestral de un experimento aleatorio e inicialmente tal conjunto no tiene ninguna estructura matemática asociada pues sus elementos pueden ser de muy diversa naturaleza. Este conjunto puede estar constituido por números (mediciones), personas, objetos, categorías, etc. Hemos señalado que no existe necesariamente un único espacio muestral para un experimento aleatorio pero en la mayoría de los casos su especificación queda entendida de manera implícita. Como hemos mencionado en la sección anterior, la  $\sigma$ -álgebra  $\mathcal{F}$  tiene el objetivo de agrupar en una sola colección a todos los subconjuntos de  $\Omega$ , llamados eventos, para los cuales uno está interesado en definir o calcular su probabilidad. Así, la  $\sigma$ -álgebra es un conjunto de subconjuntos de  $\Omega$  y existen varias  $\sigma$ -álgebras que uno puede asociar a un mismo espacio muestral.

Finalmente, la medida de probabilidad  $P$  es una función definida sobre la  $\sigma$ -álgebra, la cual no se especifica de manera totalmente explícita pero se le pide que cumpla los axiomas de Kolmogorov. Así, no existe siempre una única medida de probabilidad asignada sino que ésta es general. Tal función indica la probabilidad de ocurrencia de cada uno de los eventos contenidos en la  $\sigma$ -álgebra  $\mathcal{F}$ , sin especificar la forma concreta en que estas probabilidades son calculadas.

Supondremos entonces que para cada experimento aleatorio existe un espacio de probabilidad asociado  $(\Omega, \mathcal{F}, P)$ , el cual no es necesariamente único. Esta terna es un modelo matemático cuyo objetivo es capturar los elementos esenciales para el estudio científico del experimento aleatorio. Para la mayoría de los experimentos aleatorios que mencionaremos no se especifica con total detalle cada uno de los elementos del espacio de probabilidad, sino que éstos son entendidos de manera implícita, con lo cual se corre el riesgo de provocar alguna confusión o aparente paradoja en determinados casos y ello no es raro que ocurra.





en una lista cada elemento de  $A$ , por ejemplo, ¿cuántos números telefónicos existen que contengan por lo menos un cinco? Es poco factible que alguien intente escribir uno a uno todos estos números telefónicos y encuentre de esta manera la cantidad buscada. En las siguientes secciones estudiaremos algunas técnicas de conteo que nos ayudarán a calcular la cardinalidad de un evento  $A$  en ciertos casos particulares. El principio de multiplicación que enunciaremos a continuación es la base de muchos de los cálculos en las técnicas de conteo.

**Proposición 1.9 (Principio de multiplicación)** Si un procedimiento  $A_1$  puede efectuarse de  $n$  formas distintas y un segundo procedimiento  $A_2$  puede realizarse de  $m$  formas diferentes, entonces el total de formas en que puede efectuarse el primer procedimiento seguido del segundo es el producto  $n \cdot m$ , es decir,

$$\#(A_1 \times A_2) = \#A_1 \cdot \#A_2.$$

Para ilustrar este resultado considere el siguiente ejemplo: suponga que un cierto experimento aleatorio consiste en lanzar un dado y después seleccionar al azar una letra del alfabeto. ¿Cuál es la cardinalidad del correspondiente espacio muestral? El experimento de lanzar un dado tiene 6 resultados posibles y consideremos que tenemos un alfabeto de 26 letras. El correspondiente espacio muestral tiene entonces cardinalidad  $6 \times 26 = 156$ . El principio de multiplicación es válido no solamente para dos procedimientos sino que también vale para cualquier sucesión finita de procedimientos. Por ejemplo, si  $A_1, A_2, \dots, A_k$  denotan  $k$  procedimientos sucesivos, entonces este principio se puede enunciar en símbolos de la forma siguiente:

$$\#(A_1 \times \dots \times A_k) = \#A_1 \cdot \dots \cdot \#A_k.$$

**Ejemplo 1.14** Un hombre tiene 4 pantalones distintos, 6 camisas, y dos pares de zapatos. ¿De cuántas formas distintas puede el hombre vestirse con estas prendas? Solución:  $4 \times 6 \times 2 = 48$ . Es decir, el hombre se puede vestir de manera distinta durante 48 días sin repetir una combinación. •

Vamos a considerar a continuación diferentes esquemas y contextos en donde es posible encontrar una fórmula matemática para ciertos problemas de conteo. En todos ellos aplicaremos el principio de multiplicación. El esquema general es el de extraer al azar  $k$  objetos, uno a la vez, de una urna con  $n$  objetos distintos. Esto se muestra en la Figura 1.14.

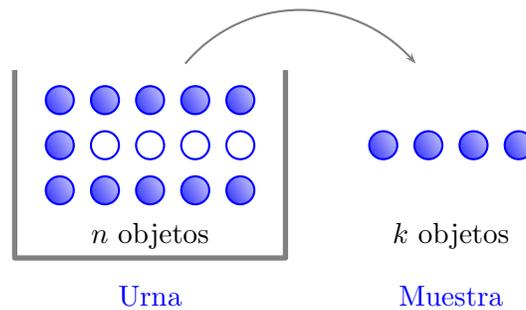


Figura 1.14

### Ordenaciones con repetición: muestras con orden y con reemplazo

Suponga que tenemos una urna con  $n$  objetos distintos. Deseamos realizar  $k$  extracciones al azar de un objeto a la vez. Al efectuar una extracción, registramos el objeto escogido y lo regresamos a la urna, de esta forma el mismo objeto puede ser extraído varias veces. El total de arreglos que se pueden obtener de esta urna al hacer  $k$  extracciones es el número  $n^k$ , pues en cada extracción tenemos  $n$  objetos posibles para escoger y efectuamos  $k$  extracciones. Esta fórmula es consecuencia del principio de multiplicación enunciado antes. A este número se le llama ordenaciones con repetición. Se dice que la muestra es con orden pues es importante el orden en el que se van obteniendo los objetos, y es con reemplazo pues cada objeto seleccionado se reincorpora a la urna.

**Ejemplo 1.15** Suponga que tenemos un conjunto de 60 caracteres diferentes. Este conjunto contiene todas las letras minúsculas del alfabeto, las letras mayúsculas, los diez dígitos y algunos caracteres especiales. ¿Cuántos *passwords* o palabras clave de longitud 4 se pueden construir usando es-

te conjunto de 60 caracteres? Este es un ejemplo de una ordenación de 60 caracteres en donde se permiten las repeticiones. Como cada caracter de los 60 disponibles puede ser escogido para ser colocado en cada una de las cuatro posiciones de la palabra clave, entonces se pueden construir  $60 \times 60 \times 60 \times 60 = 60^4 = 12,960,000$  distintos *passwords* de longitud 4. •

### Ordenaciones sin repetición: muestras con orden y sin reemplazo

Suponga que se tiene la misma situación que antes, una urna con  $n$  objetos y de los cuales se deben extraer, uno a uno,  $k$  objetos. Suponga esta vez que el muestreo es sin reemplazo, es decir, una vez seleccionado un objeto éste ya no se reincorpora a la urna. El total de arreglos distintos que se pueden obtener de este modo es el número:  $n(n-1)(n-2) \cdots (n-k+1)$ . Primeramente debemos observar que hay  $k$  factores en la expresión anterior. El primer factor es  $n$  y ello es debido a que tenemos cualesquiera de los  $n$  objetos para ser colocado en primera posición, para la segunda posición tenemos ahora  $n-1$  objetos, para la tercera  $n-2$  objetos, y así sucesivamente. Este razonamiento termina al escoger el  $k$ -ésimo objeto para el cual tenemos únicamente  $n-k+1$  posibilidades. Nuevamente por el principio de multiplicación, la respuesta es el producto indicado. La expresión encontrada puede escribirse como sigue:

$$P(n, k) = \frac{n!}{(n-k)!},$$

y se le llama permutaciones de  $n$  en  $k$ . En el caso particular cuando la muestra es exhaustiva, es decir, cuando  $k = n$ , o bien cuando todos los objetos son extraídos uno por uno, entonces se tienen todas las permutaciones o distintos órdenes en que se pueden colocar  $n$  objetos, es decir,  $n!$

**Ejemplo 1.16** ¿De cuantas formas distintas pueden asignarse los premios primero, segundo y tercero en una rifa de 10 boletos numerados del 1 al 10? Claramente se trata de una ordenación sin repetición de 10 objetos en donde se deben extraer 3 de ellos. La respuesta es entonces que existen  $10 \times 9 \times 8 = 720$  distintas asignaciones para los tres primeros lugares en la rifa. •

### Permutaciones: muestras exhaustivas con orden y sin reemplazo

La pregunta básica acerca del total de formas en que podemos poner en orden lineal (uno detrás de otro y por lo tanto no hay repetición)  $n$  objetos distintos tiene como respuesta el factorial de  $n$ , denotado por  $n!$  y definido como sigue:

$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

A este número también se le conoce como las permutaciones de  $n$  objetos, y se usa la notación  $P(n) = n!$ . Adicionalmente y por conveniencia se define  $0! = 1$ . Observe que las permutaciones de  $n$  objetos es un caso particular de la situación mencionada en la sección anterior sobre ordenaciones sin repetición cuando la muestra es exhaustiva, es decir, cuando se extraen uno a uno todos los objetos de la urna.

**Ejemplo 1.17** Si deseamos conocer el total de formas distintas en que podemos colocar una enciclopedia de 5 volúmenes en un librero, la respuesta es claramente  $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$ . El razonamiento es el siguiente: Cualquiera de los cinco libros puede ser colocado al principio, quedan cuatro libros por colocar en la segunda posición, restan entonces tres posibilidades para la tercera posición, etc. Por el principio de multiplicación la respuesta es el producto de estos números. •

### Combinaciones: muestras sin orden y sin reemplazo

Supongamos nuevamente que tenemos un conjunto de  $n$  objetos distinguibles y nos interesa obtener una muestra de tamaño  $k$ . Supongamos ahora que las muestras deben ser sin orden y sin reemplazo. Es decir, en la muestra no debe haber elementos repetidos, pues no hay reemplazo, y además la muestra debe verse como un conjunto pues no debe haber orden entre sus elementos. ¿Cuántas diferentes muestras podemos obtener de estas características? Para responder a esta pregunta seguiremos el siguiente razonamiento: cuando el orden importa hemos encontrado antes la fórmula

$$\frac{n!}{(n-k)!}.$$

Ahora que no nos interesa el orden, observamos que cada uno de los arreglos de la fórmula anterior, está siendo contado  $k!$  veces, las veces en que los

mismos  $k$  elementos pueden ser permutados unos con otros, siendo que el conjunto de elementos es el mismo. Para obtener arreglos en donde el orden no importa, debemos entonces dividir por  $k!$ . La fórmula a la que hemos llegado se llama combinaciones de  $n$  en  $k$  y la denotaremos como sigue:

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

A este número también se le conoce con el nombre de coeficiente binomial de  $n$  en  $k$ , pues aparece en el famoso teorema del binomio: para cualesquiera números reales  $a$  y  $b$ , y para cualquier número entero  $n \geq 0$ ,

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k.$$

Para los casos  $n = 2$  y  $n = 3$  el teorema del binomio se reduce a las siguientes fórmulas que muy seguramente el lector conoce:

$$\begin{aligned} (a+b)^2 &= a^2 + 2ab + b^2. \\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3. \end{aligned}$$

**Ejemplo 1.18** ¿Cuántos equipos distintos de tres personas pueden escogerse de un grupo de 5 personas?

Respuesta: observe que el orden de las tres personas escogidas no es importante de modo que la solución es

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10.$$

•

El coeficiente binomial es también una forma de generar las entradas del así llamado triángulo de Pascal, que puede observarse en Figura 1.15. El  $n$ -ésimo renglón del triángulo de Pascal, iniciando desde cero, contiene los coeficientes del desarrollo de  $(a+b)^n$ . Existe una forma sencilla de construir este triángulo observando que cada uno de estos números, exceptuando los extremos, es la suma de los dos números inmediatos del renglón anterior. A este respecto véase por ejemplo el Ejercicio 75 en la página 61.

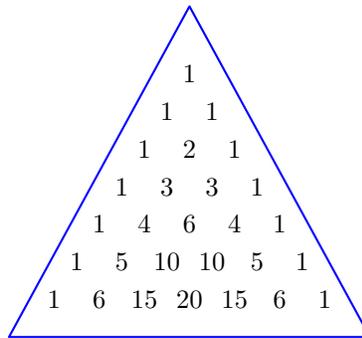


Figura 1.15: Triángulo de Pascal.

### Coefficiente multinomial

Ahora consideremos que tenemos  $n$  objetos no necesariamente distintos unos de otros, por ejemplo, supongamos que tenemos  $k_1$  objetos de un primer tipo,  $k_2$  objetos de un segundo tipo, y así sucesivamente, hasta  $k_m$  objetos del tipo  $m$ , en donde  $k_1 + k_2 + \cdots + k_m = n$ . Estos  $n$  objetos pueden todos ordenarse uno detrás de otro de tantas formas distintas como indica el así llamado coeficiente multinomial:

$$\binom{n}{k_1 k_2 \cdots k_{m-1} k_m} = \frac{n!}{k_1! k_2! \cdots k_{m-1}! k_m!}.$$

Un razonamiento para obtener esta fórmula es el siguiente: si consideramos que los  $n$  objetos son todos distintos, entonces claramente las distintas formas en que pueden escribirse todos estos objetos uno detrás de otro es  $n!$ . Pero para cada uno de estos arreglos, los  $k_1$  objetos del primer tipo, supuestos inicialmente distintos cuando en realidad no lo son, pueden permutarse entre sí de  $k_1!$  formas diferentes, siendo que el arreglo total es el mismo. De aquí que debemos dividir por  $k_1!$ . Lo mismo sucede con los elementos del segundo tipo y así sucesivamente hasta los elementos del tipo  $m$ .

**Ejemplo 1.19** ¿Cuántas palabras distintas se pueden formar permutando las letras de la palabra “mama”?

Respuesta: Existen  $\binom{4}{2,2} = 6$  palabras distintas y éstas son:

mama      amma      mmaa

maam      amam      aamm.

•

El coeficiente multinomial aparece en la siguiente fórmula:

$$(a_1 + a_2 + \cdots + a_m)^n = \sum \binom{n}{k_1 \cdots k_m} a_1^{k_1} a_2^{k_2} \cdots a_m^{k_m}, \quad (1.1)$$

en donde la suma se efectúa sobre todos los posibles valores enteros no negativos de  $k_1, k_2, \dots, k_m$ , tales que  $k_1 + k_2 + \cdots + k_m = n$ . A este resultado se le conoce como el teorema multinomial y es claramente una extensión del teorema del binomio. Por ejemplo, compruebe el lector que la fórmula (1.1) produce la siguiente expresión:

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$

¿Puede usted desarrollar  $(a + b + c)^3$ ? Es interesante observar que cuando hay únicamente dos tipos de objetos, el coeficiente multinomial se reduce al coeficiente binomial.

### Muestras sin orden y con reemplazo

Finalmente consideremos el caso de hacer  $k$  extracciones de una urna de  $n$  objetos con las condiciones de que cada objeto extraído es regresado a la urna (y entonces puede ser elegido nuevamente), y en donde el orden de la muestra no es relevante. Para encontrar una fórmula para el total de muestras que pueden obtenerse con estas características usaremos una modelación distinta pero equivalente.

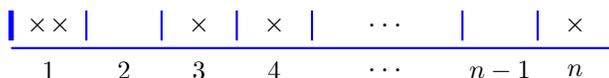


Figura 1.16

Consideremos el arreglo de  $n$  casillas de la Figura 1.16 junto con la siguiente interpretación: la primera casilla tiene dos cruces y eso indica que la bola uno

fue seleccionada dos veces, la segunda casilla esta vacía y ello significa que la bola dos no fue seleccionada, etc. El número de cruces en la casilla  $i$  indica entonces el número de veces que la bola  $i$  fue seleccionada. En total debe haber  $k$  cruces pues es el total de extracciones. Deseamos entonces conocer el número de posibles arreglos que pueden obtenerse con estas características, y debe ser claro, después de algunos momentos de reflexión, que éste es el número de muestras de tamaño  $k$ , con reemplazo y sin orden, que se pueden obtener de un conjunto de  $n$  elementos distinguibles. Consideremos que las dos paredes en los extremos de este arreglo son fijas, estas paredes se encuentran ligeramente remarcadas como puede apreciarse en la Figura 1.16. Consideremos además que las posiciones intermedias, cruz o línea vertical, pueden moverse. En total hay  $n + k - 1$  objetos movibles y cambiar de posición estos objetos produce las distintas configuraciones posibles que nos interesan. El número total de estos arreglos es entonces

$$\binom{n + k - 1}{k}$$

que equivale a colocar dentro de las  $n + k - 1$  posiciones las  $k$  cruces, dejando en los lugares restantes las paredes movibles.

### Resumen de fórmulas

En el contexto de muestras de tamaño  $k$  tomadas de un conjunto de cardinalidad  $n$ , y a manera de resumen parcial, tenemos la tabla de fórmulas en la Figura 1.17.

Muestras	con reemplazo	sin reemplazo
con orden	$n^k$	$\frac{n!}{(n - k)!}$
sin orden	$\binom{n + k - 1}{k}$	$\binom{n}{k}$

Figura 1.17

Debemos hacer énfasis, sin embargo, en que los problemas de conteo pueden ser difíciles de resolver y que para resolver un problema en particular

no debemos clasificarlo forzosamente y de manera mecánica en alguno de los esquemas mencionados. Muy posiblemente el problema en cuestión requerirá de un razonamiento especial que involucre alguna combinación de las fórmulas encontradas. En algunos casos uno puede encontrar dos o más “soluciones” distintas y aparentemente correctas de un problema. A veces las múltiples soluciones se deben a que el problema no está bien especificado y por lo tanto pueden surgir ambigüedades en su interpretación. Véase el libro de Székely [19] para conocer una amplia gama de paradojas que surgen en la probabilidad y la estadística.

Otro aspecto que amerita mencionarse es que usaremos las técnicas de conteo principalmente para aplicarlas en la definición clásica de la probabilidad:  $P(A) = \#A/\#\Omega$ . Aunque estos métodos son bastante útiles, constituyen sólo una parte mínima de la teoría de la probabilidad. Mencionaremos también que la programación de computadoras puede ser una herramienta útil para resolver problemas de conteo o simplemente para verificar la respuesta encontrada.

## Ejercicios

61. Un dado equilibrado se lanza 6 veces consecutivas. ¿Cuál es la probabilidad de que:

- a) aparezcan las seis caras del dado en orden creciente o decreciente?
- b) aparezcan las seis caras del dado en cualquier orden?
- c) sólo aparezcan números pares?
- d) aparezcan números pares e impares alternados?

Respuestas: a)  $\frac{1}{3 \cdot 6^5}$ . b)  $\frac{5}{3^2 6^2}$ . c)  $\frac{1}{2^6}$ . d)  $\frac{1}{2^5}$ .

62. **El problema de los cumpleaños.** Calcule la probabilidad de que en un grupo de  $n$  personas al menos dos de ellas tengan la misma fecha de cumpleaños.

$$\text{Respuesta: } \begin{cases} 1 - \frac{365!}{(365-n)!(365)^n} & \text{si } 2 \leq n \leq 365, \\ 1 & \text{si } n \geq 366. \end{cases}$$

63. Sea  $n \geq$  un entero. Se escogen  $n$  puntos al azar en el intervalo  $[0, L]$  de manera independiente uno de otro. Se divide el segmento en  $n$  partes de idéntica longitud. Calcule la probabilidad de que exactamente un punto caiga en cada subintervalo.

Respuesta:  $\frac{n!}{n^n}$ .

64. Se lanza un dado equilibrado tres veces. Calcule la probabilidad de obtener tres números en orden ascendente no necesariamente consecutivos.

Respuesta:  $5/54$ .

65. Sean  $A$ ,  $B$  y  $C$  tres eventos de un experimento aleatorio cuyas intersecciones son no vacías. Determine el número máximo de formas distintas en las que el evento  $A \cup B \cup C$  puede expresarse como la unión de tres eventos disjuntos. Suponga que la descomposición admite el conjunto vacío como alguno de sus tres componentes.

Respuesta: 300.

66. Suponga que se desea dibujar en un diagrama de Venn el evento  $A_1 \cup A_2 \cup \dots \cup A_n$ . Determine el número máximo de regiones simples disjuntas de las que consta este evento.

Respuesta:  $2^n - 1$ .

67. ¿De cuántas maneras diferentes pueden clasificarse los tres primeros lugares de una carrera de  $n$  corredores?

Respuesta:  $\frac{n!}{(n-3)!}$ .

68. ¿De cuántas maneras diferentes pueden sentarse  $n$  personas en una mesa circular?

Respuesta:  $(n - 1)!$

69. Suponga que los conductores de 8 automóviles estacionan sus coches completamente al azar en un estacionamiento de 12 lugares y que la configuración del estacionamiento es lineal. Determine la probabilidad de que los lugares no ocupados sean adyacentes.

Respuesta:  $1/55$ .

70. **Rumores.** En un pueblo de  $n+1$  habitantes uno de ellos le rumorea algo a una segunda persona, ésta a su vez se lo cuenta a una tercera persona

(que puede ser la primera persona) y así sucesivamente. Determine la probabilidad de que el rumor se transmita  $r$  veces sin que regrese a la primera persona.

Respuesta:  $\frac{n(n-1)^{r-1}}{n^r}$ .

71. Calcule la probabilidad de que la suma de los resultados de lanzar dos dados equilibrados sea 8 suponiendo que:

- a) los dados son distinguibles.
- b) los dados son indistinguibles.

Respuestas: a)  $1/9$ . b)  $1/7$ .

72. **Funciones.** Sean  $A$  y  $B$  dos conjuntos finitos con cardinalidades  $\#A = n$  y  $\#B = m$ . Determine el número total de funciones

$$f : A \rightarrow B$$

que sean:

- a) sin restricción alguna.
- b) inyectivas (uno a uno), suponiendo  $n \leq m$ .
- c) suprayectivas (sobre), suponiendo  $n \geq m$ .

$$\text{Respuestas: } \begin{cases} a) m^n. \\ b) m!/(m-n)! \\ c) \sum \binom{n}{k_1!k_2! \cdots k_m!} \end{cases}$$

73. Un panadero elabora 100 panes en un día en donde 10 de ellos pesan menos de lo que deberían. Un inspector pesa 5 panes tomados al azar. Calcule la probabilidad de que el inspector encuentre en su muestra exactamente un pan de peso incorrecto.

Respuesta:  $\frac{\binom{10}{1}\binom{90}{4}}{\binom{100}{5}}$ .

74. Demuestre que

$$\sum_{i=0}^k \binom{n}{i} \binom{m}{k-i} = \binom{n+m}{k}.$$

Sugerencia:  $(1+x)^n(1+x)^m = (1+x)^{n+m}$ .

75. Demuestre que

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

A partir de esta fórmula se construye el triángulo de Pascal. Más generalmente demuestre que

$$\begin{aligned} \binom{n}{k_1 k_2 \cdots k_m} &= \binom{n-1}{k_1-1 k_2 \cdots k_m} + \binom{n-1}{k_1 k_2-1 \cdots k_m} \\ &\quad + \cdots + \binom{n-1}{k_1 k_2 \cdots k_m-1}, \end{aligned}$$

en donde

$$\binom{n}{k_1 k_2 \cdots k_m} = \frac{n!}{k_1! k_2! \cdots k_m!}.$$

76. Use el ejercicio anterior y el método de inducción sobre  $n$  para demostrar la siguiente extensión del teorema del binomio:

$$(x_1 + \cdots + x_m)^n = \sum \binom{n}{k_1 k_2 \cdots k_m} x_1^{k_1} \cdots x_m^{k_m},$$

en donde la suma se realiza sobre todos los posibles valores del vector  $(k_1, \dots, k_m)$  tal que

i)  $0 \leq k_i \leq n, \quad i = 1, \dots, m.$

ii)  $k_1 + \cdots + k_m = n.$

77. Sea  $n \geq 2$  un número natural fijo. Determine la cantidad de parejas  $(x, y)$  que existen de números naturales  $x$  y  $y$  tales que

$$1 \leq x \leq y \leq n.$$

Respuesta:  $\frac{n(n+1)}{2}$ .

78. Sean  $1 \leq k \leq n$  dos números naturales. Determine la cantidad de parejas de números naturales  $(x, y)$  que existen de tal forma que

$$1 \leq x, y \leq n \quad \text{y} \quad |x - y| \geq k.$$

Respuesta:  $(n - k)(n - k + 1)$ .

79. Sean  $1 \leq k \leq n$  dos números naturales. Determine la cantidad de vectores que existen de la forma  $(x_1, \dots, x_k)$  de tal manera que cada entrada de este vector es un número entero  $x_i$  que satisface  $1 \leq x_i \leq n$  y

a)  $1 \leq x_1 < \dots < x_k \leq n$ .

b)  $1 \leq x_1 \leq \dots \leq x_k \leq n$ .

Respuestas: a)  $\binom{n}{k}$ . b)  $\binom{n+k-1}{k}$ .

80. **Las cajas de cerillos de Banach.** Una persona tiene dos cajas de cerillos. Cada caja contiene  $n$  cerillos. Coloca una caja en cada uno de sus bolsillos izquierdo y derecho, y cada vez que esta persona requiere un cerillo elije una de sus bolsillos al azar y toma de la caja correspondiente uno de los cerillos. ¿Cuál es la probabilidad de que en el momento en el que la persona se da cuenta de que la caja del bolsillo izquierdo está vacía en la caja del bolsillo derecho haya exactamente  $r$  cerillos?

Respuesta:  $\binom{2n-r}{n-r} (1/2)^{n+1} (1/2)^{n-r}$ .

81. Se lanzan tres dados equilibrados. Encuentre la probabilidad de que exactamente en dos de ellos aparezca la cara "1" suponiendo que:

a) los dados son distinguibles.

b) los dados son indistinguibles.

82. Se lanzan dos dados idénticos y equilibrados a un mismo tiempo. Calcule la probabilidad de que la suma de las dos caras sea igual a  $2, 3, \dots, 12$ . Compruebe que la suma de todas estas probabilidades es 1.

83. **Zapatos.** Una mujer tiene  $n$  pares de zapatos en desorden y en un viaje intempestivo escoge al azar  $2r$  zapatos ( $2r \leq 2n$ ). Calcule la probabilidad de que en el conjunto escogido:

a) no haya ningún par completo.

b) haya exactamente un par completo.

c) haya  $r$  pares completos.

84. **Llaves.** Una persona tiene  $n$  llaves de las cuales únicamente una ajusta a la cerradura pero no sabe cuál de ellas es la correcta. Procede a tomar las llaves al azar una por una hasta encontrar la correcta. Calcule la probabilidad de encontrar la llave correcta en el  $n$ -ésimo intento suponiendo que:

- a) retira las llaves que no funcionaron.
- b) no retira las llaves que no funcionaron.

## 1.13. Probabilidad condicional

La probabilidad condicional es un concepto elemental pero muy importante que se utiliza con mucha frecuencia en el cálculo de probabilidades. En los resultados que veremos en esta sección mostraremos las situaciones en las que se aplica la probabilidad condicional para reducir ciertas probabilidades a expresiones más sencillas.

**Definición 1.10** Sean  $A$  y  $B$  dos eventos y supongamos que  $B$  tiene probabilidad estrictamente positiva. La **probabilidad condicional** del evento  $A$  dado el evento  $B$  se denota por el símbolo  $P(A|B)$  y se define como el siguiente cociente:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1.2)$$

El término  $P(A|B)$  se lee “probabilidad de  $A$  dado  $B$ ” y es claro a partir de la definición que es necesaria la condición  $P(B) > 0$  para que el cociente esté bien definido. Cuando  $P(B) = 0$  no existe una definición establecida para  $P(A|B)$ . En ocasiones se usa la expresión  $P_B(A)$  para denotar a esta probabilidad. En la expresión (1.2), el evento  $B$  representa un evento que

ha ocurrido y la probabilidad condicional  $P(A|B)$  es la probabilidad de  $A$  modificada con la información adicional de que  $B$  ha ocurrido.

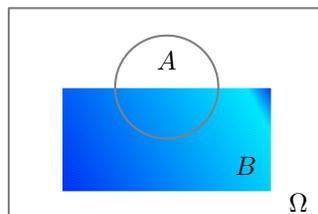


Figura 1.18

Así, uno puede imaginar que el espacio muestral del experimento aleatorio se ha reducido al evento  $B$  de tal forma que todo lo que se encuentre fuera de este evento tiene probabilidad condicional cero. La afirmación anterior es evidente a partir de observar que si  $A$  y  $B$  son ajenos, entonces el numerador de la probabilidad condicional (1.2) es cero.

**Ejemplo 1.20** Considere el experimento de lanzar un dado equilibrado y defina los eventos:

$$A = \{2\} = \text{“Cae el número 2”},$$

$$B = \{2, 4, 6\} = \text{“Cae un número par”}.$$

Es claro que  $P(A) = 1/6$ , sin embargo, sabiendo que  $B$  ha ocurrido, es decir, sabiendo que el resultado es un número par, la probabilidad del evento  $A$  es ahora

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2, 4, 6\})} = \frac{1/6}{3/6} = \frac{1}{3}.$$

•

Es inmediato verificar que la probabilidad condicional  $P(A|B)$ , vista como una función del evento  $A$ , cumple los tres axiomas de Kolmogorov, es decir, satisface:

a)  $P(\Omega|B) = 1.$

$$b) P(A|B) \geq 0.$$

$$c) P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B), \quad \text{cuando } A_1 \cap A_2 = \emptyset.$$

En consecuencia la función  $A \mapsto P(A|B)$  es una medida de probabilidad y por lo tanto se cumplen todos los resultados conocidos para cualquier medida de probabilidad, por ejemplo:

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

$$2. P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B) - P(A_1 \cap A_2 | B).$$

## Ejercicios

85. Sean  $A$  y  $B$  dos eventos tales que  $P(A) = 1/4$ ,  $P(B|A) = 1/2$  y  $P(A|B) = 1/2$ . Determine si las siguientes afirmaciones son ciertas o falsas:

a)  $A$  y  $B$  son ajenos.

$$b) P(A^c|B^c) = 3/4.$$

$$c) P(A|B) + P(A|B^c) = 1.$$

86. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

$$a) P(A|B) = P(B|A).$$

$$b) P(A|B) + P(A|B^c) = 1.$$

$$c) P(A|B) \geq P(A).$$

$$d) \text{ Si } P(A|B) \geq P(A) \text{ entonces } P(B|A) \leq P(B).$$

$$e) \text{ Si } P(A) > P(B) \text{ entonces } P(A|C) > P(B|C).$$

$$f) \text{ Si } P(A) > 0 \text{ y } P(B) > 0 \text{ entonces } P(A|B) > 0.$$

$$g) P(A) = P(B) \Leftrightarrow P(A|C) = P(B|C).$$

$$h) A \subseteq B^c \Leftrightarrow P(A|B) = 0.$$

$$i) P(A) \leq P(B) \Leftrightarrow P(A|C) \leq P(B|C).$$

$$j) A \subseteq B \Leftrightarrow P(A|C) \leq P(B|C).$$

$$k) P(A|B) = P(A|B^c) \Leftrightarrow P(B|A) = P(B|A^c).$$

$$l) \text{ Si } B \cap C = \emptyset \text{ entonces } P(A|B \cup C) = P(A|B) + P(A|C).$$

87. Para cada inciso proporcione un ejemplo en el que se cumpla la afirmación indicada. Estos ejemplos no demuestran la validez general de estas afirmaciones.

$$a) P(A|B) = 0 \text{ pero } P(A) > 0.$$

$$b) P(A|B^c) = P(A^c|B).$$

$$c) P(A) < P(A|B).$$

$$d) P(A) = P(A|B).$$

$$e) P(A) > P(A|B).$$

88. Un grupo de personas esta compuesto de 60% hombres y 40% de mujeres. De los hombres, el 30% fuma y de las mujeres el 20% fuma. Si una persona de este grupo se escoge al azar, encuentre la probabilidad de que:

a) sea hombre y fume.

b) sea hombre y no fume.

c) sea mujer y fume.

d) sea mujer y no fume.

e) sea hombre dado que se sabe que fuma.

f) sea mujer dado que se sabe que no fuma.

89. Un dado equilibrado se lanza dos veces consecutivas. Dado que en el primer lanzamiento se obtuvo un 3, ¿cuál es la probabilidad de que la suma de los dos resultados sea mayor a 6?

Respuesta: 1/2.

90. Sean  $A$  y  $B$  eventos independientes ambos con probabilidad estrictamente positiva. Demuestre que para cualquier evento  $C$ ,

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

91. Sean  $B_1, \dots, B_n$  eventos disjuntos cada uno con probabilidad estrictamente positiva. Defina el evento  $B$  como  $\bigcup_{i=1}^n B_i$  y sea  $A$  un evento tal que  $P(A | B_i) = p$  para  $i = 1, \dots, n$ . Demuestre que

$$P(A | B) = p.$$

92. **La urna de Polya.** En una urna se tienen  $r$  bolas rojas y  $b$  bolas blancas. Un ensayo consiste en tomar una bola al azar y regresarla a la urna junto con  $k$  bolas del mismo color. Se repite este ensayo varias veces y se define el evento  $R_n$  como aquél en el que se obtiene una bola roja en la  $n$ -ésima extracción. Demuestre por inducción sobre  $n$  que

$$P(R_n) = \frac{r}{r+b}, \quad n = 1, 2, \dots$$

93. **El problema de las tres puertas (Monty Hall).** Se le presentan a un concursante tres puertas cerradas detrás de una de las cuales hay un premio. El concursante debe adivinar la puerta que contiene el premio para ganarlo. Una vez que el concursante elige una puerta, y antes de abrirla, el presentador del concurso abre alguna de las puertas restantes y de la cual sabe que no contiene ningún premio. Entonces le pregunta al concursante si desea cambiar su decisión. ¿Qué debe hacer el concursante? Justifique su respuesta.
94. **El problema de los tres prisioneros.** A tres prisioneros les informa su celador que se ha escogido al azar a uno de ellos para ser ejecutado dejando a los otros dos en libertad. Uno de los prisioneros razona que tiene  $1/3$  de probabilidad de ser ejecutado y le pide al celador que le diga en secreto cuál de sus dos compañeros saldrá en libertad argumentando que por lo menos uno de ellos saldrá en libertad y que saber esta información no cambia su probabilidad de ser ejecutado. El celador, por el contrario, piensa que si el prisionero sabe cuál de sus compañeros saldrá en libertad la probabilidad de ser ejecutado aumenta a  $1/2$ . ¿Quién tiene la razón? Justifique su respuesta.
95. **El problema de la ruina del jugador.** Dos jugadores  $A$  y  $B$  lanzan sucesivamente una moneda. En cada lanzamiento, si la moneda cae cara, el jugador  $B$  le entrega una unidad monetaria al jugador  $A$ , en

caso contrario, si la moneda cae cruz, el jugador  $A$  le paga una unidad monetaria al jugador  $B$ . El juego continúa hasta que uno de ellos se arruina. Suponga que los lanzamientos de la moneda son independientes, que en cada uno de ellos la probabilidad de obtener cara es  $p$ , y que el jugador  $A$  inicia con  $n$  unidades monetarias y  $B$  inicia con  $N - n$  unidades monetarias. Defina el evento  $E_n$  como aquel en el que el jugador  $A$  gana eventualmente todo el dinero cuando comienza con  $n$  unidades monetarias y sea  $q = 1 - p$ . Demuestre que la probabilidad  $P(E_n)$ , denotada por  $p_n$ , satisface la siguiente ecuación en diferencias con las condiciones de frontera especificadas:

$$\begin{aligned} (p_{n+1} - p_n) &= \frac{q}{p} (p_n - p_{n-1}), & n = 1, 2, \dots, N - 1, \\ p_0 &= 0, \\ p_N &= 1. \end{aligned}$$

Verifique además que la solución a este sistema de ecuaciones es el siguiente:

$$p_n = \begin{cases} n/N & \text{si } p = 1/2, \\ \frac{1 - (q/p)^n}{1 - (q/p)^N} & \text{si } p \neq 1/2. \end{cases}$$

### 1.14. Teorema de probabilidad total

Sea  $\Omega$  el espacio muestral de un experimento aleatorio. Decimos que la colección de eventos  $\{B_1, B_2, \dots, B_n\}$  es una **partición finita** de  $\Omega$  si se cumplen las siguientes condiciones:

- a)  $B_i \neq \emptyset$ ,  $i = 1, 2, \dots, n$ .
- b)  $B_i \cap B_j = \emptyset$ , para  $i \neq j$ .
- c)  $\bigcup_{i=1}^n B_i = \Omega$ .

Así, se requiere que cada uno de los elementos de una partición sea distinto del conjunto vacío, que sean ajenos dos a dos y que la unión de todos ellos

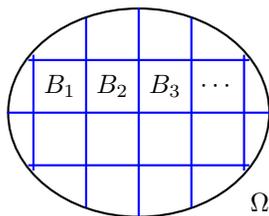


Figura 1.19

constituyan la totalidad del espacio muestral. De manera gráfica podemos representar una partición finita como se muestra en la Figura 1.19.

**Teorema 1.1 (Teorema de probabilidad total)**

Sea  $B_1, \dots, B_n$  una partición de  $\Omega$  tal que  $P(B_i) \neq 0$ ,  $i = 1, \dots, n$ . Para cualquier evento  $A$ ,

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

**Demostración.** Cualquier evento  $A$  admite la descomposición disjunta

$$A = A \cap \Omega = A \cap \left( \bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i).$$

De donde se obtiene

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

•

Cuando la partición del espacio muestral consta de únicamente los elementos  $B$  y  $B^c$ , la fórmula del teorema de probabilidad total se reduce a la expresión:

$$P(A) = P(A | B) P(B) + P(A | B^c) P(B^c).$$

En el Ejercicio 135 extenderemos ligeramente el teorema de probabilidad total al caso cuando la partición del espacio muestral consta de un número infinito numerable de elementos. La expresión es análoga:

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

Veremos algunos ejemplos de aplicación de la fórmula de probabilidad total.

**Ejemplo 1.21** Suponga que tenemos dos cajas: una con 3 bolas blancas y 7 bolas de color gris, la otra con 6 blancas y 6 grises. Esta situación se ilustra en la Figura 1.20. Si se elije una caja al azar y después se saca una bola, ¿cuál es la probabilidad de que sea blanca?

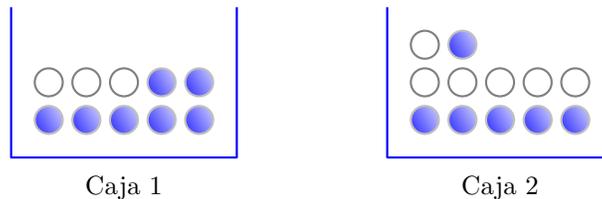


Figura 1.20

El experimento aleatorio consiste entonces en escoger una caja al azar, con idéntica probabilidad cada una de ellas, y después escoger una bola de la caja escogida. Es claro entonces que el espacio muestral puede escribirse como sigue

$$\Omega = \{(C_1, B), (C_1, G), (C_2, B), (C_2, G)\},$$

en donde  $C_1$  y  $C_2$  denotan los eventos en donde las cajas uno y dos fueron escogidas, respectivamente, y  $B$  y  $G$  denotan los eventos en donde una bola blanca o gris fueron escogidas respectivamente. Nos piden calcular la probabilidad de  $B$ . Observe que es fácil calcular la probabilidad de este evento cuando se conoce la caja que fue escogida. Esto sugiere condicionar sobre el resultado de escoger alguna de las dos cajas y aplicar el teorema de

probabilidad total, es decir,

$$\begin{aligned}P(B) &= P(B|C_1)P(C_1) + P(B|C_2)P(C_2) \\ &= \frac{3}{10} \times \frac{1}{2} + \frac{6}{12} \times \frac{1}{2} \\ &= \frac{2}{5}.\end{aligned}$$

Observe además que la partición del espacio muestral consta de dos elementos:  $\{(C_1, B), (C_1, G)\}$  y  $\{(C_2, B), (C_2, G)\}$ . ¿Puede usted comprobar que  $P(G) = 3/5$ ? Puede uno también preguntarse por situaciones aparentemente extrañas como la siguiente: si se obtuvo una bola blanca, ¿cuál es la probabilidad de que haya sido obtenida de la primera caja? Es posible calcular esta probabilidad a través del teorema de Bayes, el cual estudiaremos en la siguiente sección. •

**Ejemplo 1.22** Suponga que en una población humana de igual número de hombres y mujeres, el 4% de hombres son daltónicos y el 1% de las mujeres son daltónicas. Una persona es elegida al azar, ¿cuál es la probabilidad de que sea daltónica? Definamos primero los eventos de interés. Sea  $M$  el evento “La persona escogida es mujer”,  $H$  el evento “La persona escogida es hombre” y  $D$  el evento “La persona escogida es daltónica”. Deseamos calcular  $P(D)$ . Por el teorema de probabilidad total,

$$\begin{aligned}P(D) &= P(D|M)P(M) + P(D|H)P(H) \\ &= \frac{1}{100} \times \frac{1}{2} + \frac{4}{100} \times \frac{1}{2} \\ &= \frac{1}{40}.\end{aligned}$$

•

## Ejercicios

96. En un grupo de personas hay  $n$  mujeres y  $m$  hombres. Se escogerán a dos personas al azar de manera secuencial y sin reemplazo. Encuentre la probabilidad de que la segunda persona escogida sea mujer.

97. La urna I contiene 2 canicas blancas y 4 rojas. La urna II contiene 1 canica blanca y 1 roja. Se toma una canica al azar sin verla de la urna I y se coloca en la urna II. Después se toma una canica al azar de la urna II. Calcule la probabilidad de que la canica seleccionada de la urna II sea roja.
98. Se tiene un arreglo lineal de tres cajas como se muestra en la Figura 1.21, en donde en cada caja hay 1 canica blanca y 1 azul. Se toma una canica al azar de la primera caja y sin verla se coloca en la segunda caja. Después se toma una canica al azar de la segunda caja y sin verla se coloca en la tercera caja. Finalmente se toma una canica al azar de la tercera caja, calcule la probabilidad de que la canica escogida sea azul.

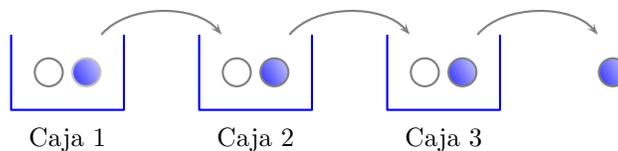


Figura 1.21

99. Se cuenta con cuatro monedas marcadas con “cara” y “cruz” tal que para la  $i$ -ésima moneda  $P(\text{“cara”}) = 0.2i$ ,  $i = 1, 2, 3, 4$ . Si se escoge una moneda al azar y se lanza al aire, encuentre la probabilidad de que ésta caiga “cruz”.
100. Una persona lanza un dado equilibrado una vez obteniendo el resultado  $x$ . Después lanza nuevamente el dado tantas veces como indicó el resultado del primer lanzamiento sumando los resultados de estos últimos lanzamientos y obteniendo un total de  $y$ . Calcule la probabilidad de que los números  $x$  y  $y$  coincidan.
101. **Canal binario ruidoso.** Se codifica un mensaje en sistema binario usando los símbolos 0 y 1. El mensaje es tal que en su codificación el 45% son ceros y el 55% son unos. Los símbolos 0 y 1 se envían uno a la vez a través de un canal binario ruidoso tal que un 0 se distorsiona en un 1 con probabilidad 0.2 y un 1 se distorsiona en un 0 con probabilidad 0.1. Encuentre la probabilidad de que en un uso cualquiera del canal:

- a) Se reciba un 0.
  - b) Se reciba un 1.
  - c) No haya error en la transmisión.
  - d) Haya algún error en la transmisión.
102. Una urna contiene 4 bolas blancas y 6 azules. Se lanza un dado equilibrado y se toma una muestra de la urna de tantas bolas como indicó el dado. Suponga que la muestra es sin reemplazo. Encuentre la probabilidad de que todas las bolas escogidas sean blancas.
103. Un estudiante contesta un examen de opción múltiple en el cual cada pregunta tiene cinco opciones como respuesta de las cuales sólo una es correcta. Cuando el estudiante conoce la respuesta correcta la selecciona, en caso contrario selecciona una de las opciones al azar. Suponga que el estudiante conoce la respuesta correcta al 80 % de las preguntas. Si el estudiante obtuvo la respuesta correcta a una pregunta, ¿cuál es la probabilidad de que el estudiante haya sabido verdaderamente la respuesta?
104. Una urna contiene 3 bola blancas y 4 negras. Se extraen dos bolas azar, una después de otra y sin reemplazo. Calcule al probabilidad de que:
- a) la segunda bola sea negra dado que la primera fue negra.
  - b) la segunda bola sea del mismo color que la primera.
  - c) la primera bola sea blanca dado que segunda fue blanca.
105. Se escogen al azar dos letras del nombre CAROLINA y se vuelven a colocar al azar en los espacios vacíos. Calcule la probabilidad de que el nombre no sea modificado.

## 1.15. Teorema de Bayes

El resultado interesante que estudiaremos a continuación involucra nuevamente probabilidades condicionales. Fue publicado por primera vez en 1763,

dos años después de la muerte de su creador: el matemático y teólogo inglés Thomas Bayes.



Thomas Bayes  
(Inglaterra 1702–1761)

**Teorema 1.2 (Teorema de Bayes)** Sea  $B_1, \dots, B_n$  una partición de  $\Omega$  tal que  $P(B_i) \neq 0$ ,  $i = 1, \dots, n$ . Sea  $A$  un evento tal que  $P(A) \neq 0$ . Entonces para cada  $j = 1, 2, \dots, n$ ,

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

**Demostración.** Por definición de probabilidad condicional y después usando el teorema de probabilidad total, tenemos que

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

•

Cuando la partición de  $\Omega$  consta de los elementos  $B$  y  $B^c$ , el teorema de Bayes para el evento  $B$  adquiere la forma:

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

Y cuando la partición consta de un número infinito numerable de elementos, el teorema de Bayes tiene la siguiente extensión ligera la cual se pide demostrar en el Ejercicio 136.

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

Veremos ahora algunos ejemplos de aplicación del teorema de Bayes.

**Ejemplo 1.23** En una fábrica hay dos máquinas. La máquina 1 realiza el 60 % de la producción total y la máquina 2 el 40 %. De su producción total, la máquina 1 produce 3 % de material defectuoso, la 2 el 5 %. El asunto es que se ha encontrado un material defectuoso, ¿cuál es la probabilidad de que este material defectuoso provenga de la máquina 2?

Solución: Sea  $M_1$  el evento “La máquina 1 produjo el material escogido”,  $M_2$  en evento “La máquina 2 produjo el material escogido” y finalmente sea  $D$  el evento “El material escogido es defectuoso”. El problema es encontrar  $P(M_2|D)$  y observamos que la información que tenemos es  $P(D|M_2)$ . Por el teorema de Bayes tenemos entonces que

$$\begin{aligned} P(M_2|D) &= \frac{P(D|M_2)P(M_2)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2)} \\ &= \frac{\frac{5}{100} \times \frac{40}{100}}{\frac{3}{100} \times \frac{60}{100} + \frac{5}{100} \times \frac{40}{100}} \\ &= \frac{10}{19}. \end{aligned}$$

Como un ejercicio al lector se deja comprobar que  $P(M_1|D) = 9/19$ . •

**Ejemplo 1.24** En un laboratorio se descubrió una prueba para detectar cierta enfermedad y sobre la eficacia de dicha prueba se conoce lo siguiente: si se denota por  $E$  el evento de que un paciente tenga la enfermedad y por  $N$  el evento de que la prueba resulte negativa, entonces se sabe que  $P(N^c | E) = 0.95$ ,  $P(N | E^c) = 0.96$  y  $P(E) = 0.01$ . Con esta información uno podría pensar que la prueba es muy buena, sin embargo calcularemos las probabilidades  $P(E | N)$  y  $P(E | N^c)$  usando el teorema de Bayes.

$$\begin{aligned} P(E | N) &= \frac{P(N | E)P(E)}{P(N | E)P(E) + P(N | E^c)P(E^c)} \\ &= \frac{0.05 \times 0.01}{0.05 \times 0.01 + 0.96 \times 0.99} \\ &= 0.000526. \end{aligned}$$

Es bueno que esta probabilidad sea pequeña, pero por otro lado,

$$\begin{aligned} P(E | N^c) &= \frac{P(N^c | E)P(E)}{P(N^c | E)P(E) + P(N^c | E^c)P(E^c)} \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.04 \times 0.99} \\ &= 0.193. \end{aligned}$$

Esta última probabilidad es demasiado pequeña y por lo tanto la prueba no es muy confiable en tales casos. •

## Ejercicios

106. **La paradoja de la caja de Bertrand.** Se tienen tres cajas y cada una de ellas tiene dos monedas. La caja  $C_1$  contiene dos monedas de oro. La caja  $C_2$  contiene una moneda de oro y una de plata. La caja  $C_3$  contiene dos monedas de plata. Véase la Figura 1.22. Se selecciona una caja al azar y de allí se escoge una moneda. Si resulta que la moneda escogida es de oro, ¿cuál es la probabilidad de que provenga de la caja con dos monedas de oro?

- a) Argumente con palabras que la respuesta es  $1/2$ .  
 b) Demuestre que la respuesta es  $2/3$ .

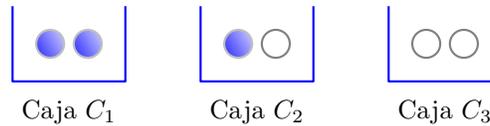


Figura 1.22

107. Una persona toma al azar uno de los números 1, 2 o 3, con idéntica probabilidad cada uno de ellos y luego tira un dado equilibrado tantas veces como indica el número escogido. Finalmente suma los resultados de las tiradas del dado. Calcule la probabilidad de que:

- a) se obtenga un total de 5.  
 b) se haya escogido el número 2 dado que la suma de las tiradas del dado es 8.

Respuestas: a)  $11/108$ .  
 b)  $1/4$ .

108. Suponga que se cuenta con dos urnas con la siguiente configuración: la urna  $A$  contiene 1 bola negra y 1 blanca, la urna  $B$  contiene 2 bolas negras y 2 blancas. Véase la Figura 1.23. Se escogen al azar dos bolas de la urna  $B$  y se transfieren a la urna  $A$ . Se escoge después una bola al azar de la urna  $A$ . Calcule la probabilidad de que:

- a) la bola escogida sea blanca.  
 b) al menos una bola blanca haya sido transferida dado que la bola escogida es blanca.



Figura 1.23

109. **Canal binario ruidoso.** Se codifica un mensaje en sistema binario usando los símbolos 0 y 1. El mensaje es tal que en su codificación el 45 % son ceros y el 55 % son unos. Los símbolos 0 y 1 se envían uno a la vez a través de un canal binario ruidoso tal que un 0 se distorsiona en un 1 con probabilidad 0.2 y un 1 se distorsiona en un 0 con probabilidad 0.1. Encuentre la probabilidad de que en un uso cualquiera del canal:
- Se haya enviado un 0 dado que se recibió un 0.
  - Se haya enviado un 1 dado que se recibió un 1.
110. Una caja contiene 3 bolas blancas y 4 bolas azules como se muestra en la Figura 1.24. Suponga que se extraen dos bolas al azar, una después de otra y sin reemplazo. Calcule la probabilidad de que:
- la segunda bola sea azul dado que la primera es azul.
  - la segunda bola sea del mismo color que la primera.
  - las dos bolas sean de distinto color.
  - la primera bola sea azul dado que la segunda es azul.
  - la segunda bola sea azul.

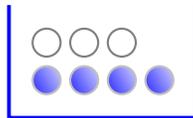


Figura 1.24

111. La urna I contiene 2 canicas blancas y 4 rojas. La urna II contiene 1 canica blanca y 1 roja. Se toma una canica al azar sin verla de la urna I y se coloca en la urna II. Después se toma una canica al azar de la urna II. Calcule la probabilidad de que la canica seleccionada de la urna I haya sido roja dado que la que se obtuvo de la urna II es roja.

## 1.16. Independencia de eventos

El concepto de independencia es una forma de incorporar al cálculo de probabilidades la no afectación de la ocurrencia de un evento sobre la probabilidad de otro. Es un concepto importante que se deriva de observaciones de situaciones reales y su utilización reduce considerablemente el cálculo de probabilidades.

**Definición 1.11** Se dice que los eventos  $A$  y  $B$  son **independientes** si se cumple la igualdad

$$P(A \cap B) = P(A)P(B). \quad (1.3)$$

Bajo la hipótesis adicional de que  $P(B) > 0$ , la identidad (1.3) puede escribirse como  $P(A|B) = P(A)$  y esto significa que la ocurrencia del evento  $B$  no afecta a la probabilidad de  $A$ . Análogamente, cuando  $P(A) > 0$  la identidad (1.3) se puede escribir como  $P(B|A) = P(B)$ , es decir, la ocurrencia del evento  $A$  no cambia a la probabilidad de  $B$ . Veamos algunos ejemplos.

**Ejemplo 1.25** Considere un experimento aleatorio con espacio muestral equiprobable  $\Omega = \{1, 2, 3, 4\}$ .

1. Los eventos  $A = \{1, 2\}$  y  $B = \{1, 3\}$  son independientes pues tanto  $P(A \cap B)$  como  $P(A)P(B)$  coinciden en el valor  $1/4$ .
2. Los eventos  $A = \{1, 2, 3\}$  y  $B = \{1, 3\}$  no son independientes pues  $P(A \cap B) = 1/2$  mientras que  $P(A)P(B) = (3/4)(1/2) = 3/8$ .

•

**Ejemplo 1.26 (Dos eventos que no son independientes).** Recordemos que un ensayo en una urna de Polya consiste en escoger una bola al azar de una urna con una configuración inicial de  $r$  bolas rojas y  $b$  bolas blancas y que la bola escogida se regresa a la urna junto con  $c$  bolas del mismo color. Considere los eventos  $R_1$  y  $R_2$  de la urna del Polya, es decir,

obtener una bola roja en la primera y en la segunda extracción, respectivamente. Es claro que estos eventos no son independientes pues

$$P(R_2 | R_1)P(R_1) \neq P(R_2).$$

•

Inicialmente uno podría asociar la idea de independencia de dos eventos con el hecho de que éstos son ajenos, pero ello es erróneo en general. A continuación ilustraremos esta situación.

**Ejemplo 1.27 (Independencia  $\Rightarrow$  Ajenos).** Considere un evento  $A \neq \emptyset$  junto con el espacio muestral  $\Omega$ . Es claro que  $A$  y  $\Omega$  son independientes pues  $P(A \cap \Omega) = P(A)P(\Omega)$ . Sin embargo,  $A \cap \Omega = A \neq \emptyset$ . Por lo tanto, el hecho de que dos eventos sean independientes no implica necesariamente que sean ajenos. •

Como complemento al ejemplo anterior, ilustraremos ahora el hecho de que dos eventos pueden ser ajenos sin ser independientes. Véase también el Ejercicio 116 para otro ejemplo de este tipo de situaciones.

**Ejemplo 1.28 (Ajenos  $\Rightarrow$  Independencia).** Considere el experimento aleatorio de lanzar un dado equilibrado y defina los eventos  $A$  como obtener un número par y  $B$  como obtener un número impar. Es claro que los eventos  $A$  y  $B$  son ajenos, sin embargo no son independientes pues  $P(A \cap B) \neq P(A)P(B)$ . Por lo tanto, dos eventos pueden ser ajenos y ello no implica necesariamente que sean independientes. •

La definición de independencia de dos eventos puede extenderse al caso de una colección finita de eventos de la siguiente forma.

**Definición 1.12** Decimos que  $n$  eventos  $A_1, A_2, \dots, A_n$  son independientes si se satisfacen todas y cada una de las condiciones siguientes:

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad i, j \text{ distintos.} \quad (1.4)$$

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} \quad (1.5)$$

$$\vdots$$

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdots P(A_n). \quad (1.6)$$

Observe que hay en total  $\binom{n}{2}$  identidades diferentes de la forma (1.4). Si se cumplen todas las identidades de esta forma, decimos que la colección de eventos es **independiente dos a dos**, es decir, independientes tomados por pares. Análogamente hay  $\binom{n}{3}$  identidades diferentes de la forma (1.5). Si se cumplen todas las identidades de esta forma, decimos que la colección de eventos es **independiente tres a tres**, es decir, independientes tomados por tercias. Y así sucesivamente hasta la identidad de la forma (1.6) de la cual sólo hay una expresión.

En general, para verificar que  $n$  eventos son independientes es necesario comprobar todas y cada una de las igualdades arriba enunciadas. Es decir, cualquiera de estas igualdades no implica, en general, la validez de alguna otra. En los ejemplos que aparecen abajo mostramos que la independencia dos a dos no implica la independencia tres a tres, ni viceversa. No es difícil darse cuenta que el total de igualdades que es necesario verificar para que  $n$  eventos sean independientes es  $2^n - n - 1$ .

**Ejemplo 1.29 (Independencia dos a dos  $\not\Rightarrow$  Independencia tres a tres).** Considere el siguiente espacio muestral equiprobable junto con los siguientes eventos:

$$\Omega = \{1, 2, 3, 4\},$$

$$A = \{1, 2\}, \quad B = \{2, 3\}, \quad C = \{2, 4\}.$$

Los eventos  $A, B$  y  $C$  no son independientes pues aunque se cumplen las igualdades  $P(A \cap B) = P(A)P(B)$ ,  $P(A \cap C) = P(A)P(C)$ , y  $P(B \cap C) =$

$P(B)P(C)$ , sucede que  $P(A \cap B \cap C) \neq P(A)P(B)P(C)$ . Esto muestra que, en general, la independencia dos a dos no implica la independencia tres a tres. •

**Ejemplo 1.30 (Independencia tres a tres  $\Rightarrow$  Independencia dos a dos).** Considere ahora el siguiente espacio muestral equiprobable junto con los siguientes eventos:

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\},$$

$$A = \{1, 2, 3, 4\}, \quad B = \{1, 5, 6, 7\}, \quad C = \{1, 2, 3, 5\}.$$

Estos eventos cumplen la condición  $P(A \cap B \cap C) = P(A)P(B)P(C)$  pero  $P(A \cap B) \neq P(A)P(B)$ . Esto muestra que, en general, la independencia tres a tres no implica la independencia dos a dos. •

Finalmente mencionaremos que el concepto de independencia puede extenderse al caso de colecciones infinitas de eventos en la siguiente forma.

**Definición 1.13** Se dice que un colección infinita de eventos es **independiente** si cualquier subcolección finita de ella lo es.

## Ejercicios

112. Sean  $A$  y  $B$  eventos tales que  $P(A) = 4/10$  y  $P(A \cup B) = 7/10$ . Encuentre la probabilidad de  $B$  suponiendo que:

- a)  $A$  y  $B$  son independientes.
- b)  $A$  y  $B$  son ajenos.
- c)  $P(A|B) = 1/2$ .

Respuestas: a)  $1/2$ , b)  $3/10$ , c)  $3/5$ .

113. Se lanza un dado equilibrado dos veces. Determine si los siguientes pares de eventos son independientes:

- a)  $A =$  “La suma de los dos resultados es 6.”  
 $B =$  “El primer resultado es 4.”
- b)  $A =$  “La suma de los dos resultados es 7.”  
 $B =$  “El segundo resultado es 4.”

114. Demuestre o proporcione un contraejemplo.

- a)  $A$  y  $\emptyset$  son independientes.
- b)  $A$  y  $\Omega$  son independientes.
- c) Si  $P(A|B) = P(A|B^c)$  entonces  $A$  y  $B$  son independientes.
- d) Si  $A$  y  $B$  son independientes entonces  $P(A|B) = P(A|B^c)$ .
- e) Si  $A$  tiene probabilidad cero o uno, entonces  $A$  es independiente de cualquier otro evento.
- f) Si  $A$  es independiente consigo mismo entonces  $P(A) = 0$ .
- g) Si  $A$  es independiente consigo mismo entonces  $P(A) = 1$ .
- h) Si  $P(A) = P(B) = P(A|B) = 1/2$  entonces  $A$  y  $B$  son independientes.

115. Demuestre que las siguientes cuatro afirmaciones son equivalentes:

- a)  $A$  y  $B$  son independientes.
- b)  $A$  y  $B^c$  son independientes.
- c)  $A^c$  y  $B$  son independientes.
- d)  $A^c$  y  $B^c$  son independientes.

116. Sea  $\Omega = \{a, b, c, d\}$  un espacio muestral equiprobable. Defina los eventos  $A = \{a, b\}$ ,  $B = \{a, c\}$  y  $C = \{a\}$ . Compruebe que los siguientes pares de eventos satisfacen las propiedades indicadas.

Eventos	Ajenos	Independientes
$A, C$	×	×
$A, B$	×	✓
$A, A^c$	✓	×
$A, \emptyset$	✓	✓

117. Sean  $A$  y  $B$  dos eventos independientes. Encuentre una expresión en términos de  $P(A)$  y  $P(B)$ , o sus complementos, para las siguientes probabilidades:

- |                      |                          |
|----------------------|--------------------------|
| a) $P(A \cup B)$ .   | d) $P(A \Delta B)$ .     |
| b) $P(A \cup B^c)$ . | e) $P(A - (A \cap B))$ . |
| c) $P(A - B)$ .      | f) $P(\Omega)$ .         |

118. Sean  $A$  y  $B$  dos eventos independientes tales que  $P(A) = p$  y  $P(B) = q$ . Calcule la probabilidad de que:

- no ocurra ninguno de estos dos eventos.
- ocurra exactamente uno de estos dos eventos.
- ocurra por lo menos uno de estos dos eventos.
- ocurra a lo sumo uno de estos dos eventos.
- ocurran los dos eventos.

119. Sean  $A$  y  $B$  dos eventos independientes. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- $A$  y  $A \cup B$  son independientes.
- $A$  y  $A \cap B$  son independientes.
- $A$  y  $B - A$  son independientes.
- $A - B$  y  $B - A$  son independientes.

120. Demuestre que las siguientes cuatro afirmaciones son equivalentes:

- $A, B$  y  $C$  son independientes.
- $A^c, B$  y  $C$  son independientes.
- $A^c, B^c$  y  $C$  son independientes.
- $A^c, B^c$  y  $C^c$  son independientes.

121. Sean  $A, B$  y  $C$  tres eventos tales que  $A$  es independiente de  $B$  y  $A$  también es independiente de  $C$ . Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- a)  $A$  y  $(B \cap C)$  son independientes.  
b)  $A$  y  $(B \cup C)$  son independientes.  
c)  $B$  y  $C$  son independientes.
122. Suponga que  $A$  es un evento independiente de  $B$ , de  $C$  y de  $(B \cup C)$ , por separado. Demuestre que  $A$  y  $(B \cap C)$  son independientes.
123. En una cierta zona geográfica estratégica se han colocado tres radares para detectar el vuelo de aviones de baja altura. Cada radar funciona de manera independiente y es capaz de detectar un avión con probabilidad 0.85. Si un avión atraviesa la zona en estudio calcule la probabilidad de que:
- a) no sea detectado por ninguno de los radares.  
b) sea detectado por lo menos por uno de los radares.  
c) sea detectado por lo menos por dos de los radares.
124. Sean  $A$ ,  $B$  y  $C$  tres eventos independientes. Demuestre directamente que los siguientes pares de eventos son independientes:
- a)  $A$  y  $B \cap C$ .      c)  $(A \cap B)$  y  $C$ .      e)  $B$  y  $(A - C)$ .  
b)  $A$  y  $B \cup C$ .      d)  $(A \cup B)$  y  $C$ .      f)  $B$  y  $(A^c \cup C^c)$ .
125. Sean  $A$ ,  $B$  y  $C$  tres eventos independientes. Encuentre una expresión en términos de  $P(A)$ ,  $P(B)$  y  $P(C)$ , o sus complementos, para las siguientes probabilidades:
- a)  $P(A \cup B \cup C)$ .      d)  $P(A - B - C)$ .      g)  $P(A \cup (B \cap C))$ .  
b)  $P(A \cup B \cup C^c)$ .      e)  $P(A \cap (B \cup C))$ .      h)  $P(A \Delta B \Delta C)$ .  
c)  $P(A^c \cap B^c \cap C^c)$ .      f)  $P(A \cap B \cap C^c)$ .      i)  $P(\Omega)$ .
126. Sean  $A$ ,  $B$  y  $C$  tres eventos independientes tales que  $P(A) = p$ ,  $P(B) = q$  y  $P(C) = r$ . Calcule la probabilidad de que:
- a) no ocurra ninguno de estos tres eventos.  
b) ocurra exactamente uno de estos tres eventos.

- c) ocurra por lo menos uno de estos tres eventos.  
 d) ocurran exactamente dos de estos tres eventos.  
 e) ocurran por lo menos dos de estos tres eventos.  
 f) ocurra a lo sumo uno de estos tres eventos.  
 g) ocurran a lo sumo dos de estos tres eventos.  
 h) ocurran los tres eventos.
127. Use el teorema del binomio para demostrar que el total de igualdades por verificar para comprobar que  $n$  eventos son independientes es  $2^n - 1 - n$ .
128. Sean  $B_1$  y  $B_2$  dos eventos cada uno de ellos independiente del evento  $A$ . Proporcione un contraejemplo para las siguientes afirmaciones:
- a)  $A$  y  $B_1 \cup B_2$  son independientes.  
 b)  $A$  y  $B_1 \cap B_2$  son independientes.  
 c)  $A$  y  $B_1 - B_2$  son independientes.  
 d)  $A$  y  $B_1 \triangle B_2$  son independientes.
129. Demuestre por inducción sobre  $n$  que si  $A_1, \dots, A_n$  son eventos independientes entonces

$$P(A_1 \cup \dots \cup A_n) = 1 - (1 - P(A_1)) \cdots (1 - P(A_n)).$$

130. **Independencia condicional.** Se dice que los eventos  $A$  y  $B$  son independientes condicionalmente al evento  $C$  si

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

suponiendo de antemano que  $P(C) > 0$ . Proporcione contraejemplos adecuados para demostrar que:

- a) la independencia de dos eventos no implica necesariamente su independencia condicional dado un tercer evento.  
 b) la independencia condicional de dos eventos dado un tercero no implica necesariamente la independencia de los dos primeros.

En general pueden darse ejemplos de las situaciones que se presentan en la siguiente tabla:

$A, B$ indep.	$A, B$ indep. $  C$
✓	✓
✓	×
×	✓
×	×

131. Un cierto componente de una máquina falla el 5% de las veces que se enciende. Para obtener una mayor confiabilidad se colocan  $n$  componentes de las mismas características en un arreglo paralelo, de tal forma que ahora el conjunto de componentes falla cuando todos fallan. Suponga que el comportamiento de cada componente es independiente uno del otro. Determine el mínimo valor de  $n$  a fin de garantizar el funcionamiento de la máquina por lo menos el 99% de las veces.
132. Dos jugadores de basketbol alternan turnos para efectuar tiros libres hasta que uno de ellos enceste. En cada intento la probabilidad de encestar es  $p$  para el primer jugador y  $q$  para el segundo jugador, siendo los resultados de los tiros independientes unos de otros. Calcule la probabilidad de encestar primero de cada uno de los jugadores.

## 1.17. Continuidad de las medidas de probabilidad

Para concluir este capítulo mencionaremos una propiedad que cumple toda medida de probabilidad la cual es ligeramente más avanzada que las que hasta ahora hemos estudiado. Aunque la demostración de este resultado no es complicada, por simplicidad en esta exposición la omitiremos para concentrarnos en su interpretación y aplicación en las siguientes secciones.

**Proposición 1.10** Sea  $A_1, A_2, \dots$  una sucesión infinita de eventos no decreciente, es decir,  $A_1 \subseteq A_2 \subseteq \dots$ . Entonces

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

Como la sucesión de eventos es no decreciente entonces es natural definir su límite como la unión de todos ellos, es decir,

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

Observe que esta unión infinita no necesariamente es la totalidad del espacio muestral  $\Omega$ , pero es un subconjunto de él. Por lo tanto el resultado anterior establece que la probabilidad del límite de la sucesión coincide con el límite de las probabilidades. Este intercambio de límite y probabilidad es la definición de **continuidad**, en este caso para medidas de probabilidad y únicamente en el caso de sucesiones monótonas no decrecientes de eventos. Este resultado puede extenderse al caso de sucesiones monótonas no crecientes, es decir, cuando  $A_1 \supseteq A_2 \supseteq \dots$ , en cuyo caso se define

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

y la afirmación de continuidad es que

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

Haremos uso de estos resultados teóricos importantes en el siguiente capítulo.

## Ejercicios

133. Suponiendo válida la Proposición 1.10, demuestre que si  $A_1, A_2, \dots$  es una sucesión infinita de eventos no creciente, es decir,  $A_1 \supseteq A_2 \supseteq \dots$

Entonces

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

134. Encuentre el límite de las siguientes sucesiones de eventos, los cuales son subconjuntos de  $\mathbb{R}$ .

a)  $A_n = (-\infty, a + 1/n)$ .

e)  $A_n = (-n, n)$ .

b)  $A_n = (-\infty, a + 1/n]$ .

f)  $A_n = (a + 1/n, \infty)$ .

c)  $A_n = (-1/n, 1/n)$ .

g)  $A_n = [a - 1/n, \infty)$ .

d)  $A_n = [0, 1/n]$ .

h)  $A_n = (a - 1/n, b + 1/n)$ .

135. **Teorema de probabilidad total extendido.** Sea  $B_1, B_2, \dots$  una partición infinita de  $\Omega$  tal que  $P(B_i) \neq 0$ ,  $i = 1, 2, \dots$ . Demuestre que para cualquier evento  $A$ ,

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

136. **Teorema de Bayes extendido.** Sea  $B_1, B_2, \dots$  una partición infinita de  $\Omega$  tal que  $P(B_i) \neq 0$ ,  $i = 1, 2, \dots$ . Sea  $A$  un evento tal que  $P(A) \neq 0$ . Entonces para cada  $j = 1, 2, \dots$ ,

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

## Pierre-Simon Laplace

P.-S. Laplace (Francia, 1749–1827) nació en el seno de una familia medianamente acomodada dedicada a la agricultura y al comercio. Su padre tuvo la intención de que Laplace siguiera una carrera eclesiástica, de modo que de pequeño asistió a una escuela local a cargo de una orden religiosa católica y a la edad de 16 años fue enviado a la Universidad de Caen para estudiar teología. En esta universidad tuvo profesores de matemáticas que despertaron vivamente su interés en esta disciplina y quienes reconocieron el potencial matemático del joven estudiante.



P.-S. Laplace

Laplace dejó la Universidad de Caen sin graduarse pero consigue una carta de presentación de sus profesores y viaja a París cuando tenía 19 años para presentarse ante d'Alembert. Después de una fría y escéptica recepción por parte de d'Alembert, Laplace logra demostrarle su capacidad de trabajo y su sorprendente habilidad natural para las matemáticas. Por medio de una carta de recomendación de d'Alembert, Laplace consigue un puesto como profesor de matemáticas en la *École Militaire* de París. Una vez que Laplace asegura un ingreso económico estable, puede entonces establecerse en París y dedicarse con mayor empeño a sus trabajos de investigación. En efecto, de 1771 a 1787 logró producir gran parte de su monumental trabajo original en astronomía matemática. Contribuyó notablemente además al desarrollo de varias áreas de la ciencia, en particular la mecánica, la probabilidad, la estadística, la física matemática, el análisis matemático y las ecuaciones diferenciales y en diferencias. La diversidad de temas de sus trabajos científicos muestra su amplia capacidad intelectual y su dominio de muchas áreas de la física y la matemática de su tiempo. En 1773, a la edad de 24 años y después de varios intentos, logra ser nombrado miembro adjunto de la prestigiosa *Académie des Sciences* de Francia, donde a lo largo de los años ocupa diversos nombramientos y desde donde hace sentir la influencia de sus opiniones y pensamientos. Es en esos años cuando Laplace consolida su fama como astrónomo y matemático. En su obra monumental titulada *Mécanique Céleste*, la cual consta de cinco volúmenes y que fue publicada

a lo largo de varios años (1799-1825) , Laplace resume y extiende los resultados de sus predecesores. En este trabajo traslada los estudios geométricos de la mecánica clásica al de una mecánica basada en el cálculo diferencial e integral, abriendo con ello nuevas perspectivas de desarrollo.

En 1785, siendo Laplace examinador de estudiantes de la Escuela de Artilleros, examinó y acreditó al joven estudiante de 16 años Napoleón Bonaparte (1769-1821). En contraparte, 14 años después en 1799, Laplace fue nombrado Ministro del Interior de Francia por parte de Napoleón Bonaparte. Sin embargo, Laplace ocupó tal cargo por únicamente seis semanas y con bastante malos resultados según un reporte de Napoleón.

En 1812 Laplace publicó un tratado titulado *Theorie Analytique des Probabilités*, en el cual establece varios resultados fundamentales de la probabilidad y la estadística. En particular, estudia el concepto de probabilidad clásica como aparece definida en este texto y analiza varias de las propiedades que hemos estudiado. La primera edición de este trabajo de Laplace fue dedicada a Napoleón Bonaparte, sin embargo, tal referencia fue omitida en ediciones subsecuentes a partir de la abdicación de éste. Este pequeño hecho muestra los cambios de postura en cuestiones políticas que Laplace tuvo y que sus amigos y colegas le recriminaron. Se cree que Laplace buscaba alejarse de la política y de los asuntos de la vida ajenos a su quehacer científico, pero ello no era fácil dado el liderazgo que como hombre de ciencia naturalmente ocupaba y a los tiempos turbulentos que le tocó vivir en Francia.

A Laplace se le considera como un hombre con una extraordinaria capacidad para las matemáticas. Sin duda fue uno de los pensadores más brillantes dentro del grupo de sus contemporáneos y su influencia y contribuciones al desarrollo de la ciencia fueron notables. En honor a ello su nombre aparece grabado en la torre Eiffel. Laplace murió en París el 5 de marzo de 1827 a la edad de 77 años.

Fuente: Archivo MacTutor, Universidad de St. Andrews [22].

## Andrey Nikolaevich Kolmogorov

A. N. Kolmogorov (Rusia, 1903–1987) quedó huérfano en el momento de nacer pues en el parto mismo murió su madre. Ella no estaba casada con el padre y éste no se ocupó del recién nacido. De su padre se conoce muy poco, se cree que fue muerto en 1919 durante la guerra civil rusa. Así, el pequeño Kolmogorov es criado por dos de sus tías en la casa de su abuelo paterno, de quien adquiere el apellido Kolmogorov. En 1910 se muda a Moscú con una de sus tías y realiza allí sus primeros estudios. Desde muy pequeño mostró inclinación por la ciencia y la literatura. Siendo adolescente diseñó varias máquinas de “movimiento perpetuo” para las cuales no era evidente descubrir el mecanismo de funcionamiento. En 1920 ingresó a la Universidad Estatal de Moscú y al Instituto Tecnológico de Química, en donde estudia diversas disciplinas además de matemáticas, entre ellas metalurgia e historia de Rusia. En 1929, a la edad de 19 años empieza a tener fama mundial al publicar un trabajo matemático en donde Kolmogorov construye una serie de Fourier que es divergente casi dondequiera. Alrededor de estas fechas es cuando decide dedicarse completamente a las matemáticas. En 1925 se gradúa en la Universidad Estatal de Moscú y en 1929 obtiene el doctorado habiendo publicado para ese entonces 18 trabajos científicos. Se incorpora a la misma universidad en 1931 como profesor. En 1933 publica su famoso libro *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Fundamentos de la Teoría de la Probabilidad) en donde establece los fundamentos axiomáticos de la teoría de la probabilidad y los cuales hemos estudiado en este texto. Así, Kolmogorov es uno de los fundadores de la teoría moderna de la probabilidad, quien también contribuyó con brillantez en muchas otras áreas de las matemáticas y la física como la topología, la lógica, los fenómenos de turbulencia, la mecánica clásica, la teoría de la información y la complejidad computacional. Es interesante mencionar que Kolmogorov tuvo particular interés en un proyecto consistente en proveer educación especial a niños sobresalientes, a quienes dedicó tiempo para crearles las condiciones materiales y de estudio para que éstos tuvieran una



A. N. Kolmogorov

educación amplia e integral. Kolmogorov recibió numerosos reconocimientos por la profundidad e importancia de sus trabajos científicos, tales reconocimientos provinieron no únicamente de Rusia, sino también de otros países y de varias universidades y academias científicas internacionales.

Fuente: Archivo MacTutor, Universidad de St. Andrews [22].



## Capítulo 2

# Variables aleatorias

En este capítulo definiremos a una variable aleatoria como una función del espacio muestral en el conjunto de números reales. Esto nos permitirá considerar que el resultado del experimento aleatorio es el número real tomado por la variable aleatoria. En consecuencia, nuestro interés en el estudio de los experimentos aleatorios se trasladará al estudio de las distintas variables aleatorias y sus características particulares.

### 2.1. Variables aleatorias

Consideremos que tenemos un experimento aleatorio cualquiera junto con un espacio de probabilidad asociado  $(\Omega, \mathcal{F}, P)$ .

**Definición 2.1** Una **variable aleatoria** es una transformación  $X$  del espacio de resultados  $\Omega$  al conjunto de números reales, esto es,

$$X : \Omega \rightarrow \mathbb{R},$$

tal que para cualquier número real  $x$ ,

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}. \quad (2.1)$$

La condición (2.1) será justificada más adelante. Supongamos entonces que se efectúa el experimento aleatorio una vez y se obtiene un resultado  $\omega$  en  $\Omega$ . Al transformar este resultado con la variable aleatoria  $X$  se obtiene un número real  $X(\omega) = x$ . Así, una variable aleatoria es una función determinista y no es variable ni aleatoria, sin embargo tales términos se justifican al considerar que los posibles resultados del experimento aleatorio son los diferentes números reales  $x$  que la función  $X$  puede tomar. De manera informal puede uno pensar también que una variable aleatoria es una pregunta o medición que se hace sobre cada uno de los resultados del experimento aleatorio y cuya respuesta es un número real, así cada resultado  $\omega$  tiene asociado un único número  $x$ . De manera gráfica se ilustra el concepto de variable aleatoria en la Figura 2.1.

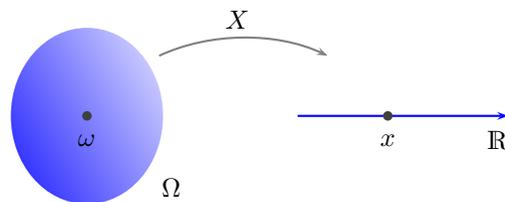


Figura 2.1

En lo sucesivo emplearemos la siguiente notación: si  $A$  es un conjunto de Borel de  $\mathbb{R}$ , entonces la expresión  $(X \in A)$ , incluyendo el paréntesis, denota el conjunto  $\{\omega \in \Omega : X(\omega) \in A\}$ , es decir,

$$(X \in A) = \{\omega \in \Omega : X(\omega) \in A\}.$$

En palabras, la expresión  $(X \in A)$  denota aquel conjunto de elementos  $\omega$  del espacio muestral  $\Omega$  tales que bajo la aplicación de la función  $X$  toman un valor dentro del conjunto  $A$ . A este conjunto se le llama la **imagen inversa** de  $A$  y se le denota por  $X^{-1}A$ , lo cual no debe confundirse con la función inversa de  $X$  pues ésta puede no existir. Véase el Ejercicio 138 en donde se pide demostrar algunas propiedades sencillas de la imagen inversa. Por

ejemplo, consideremos que el conjunto  $A$  es el intervalo  $(a, b)$ , entonces el evento  $(X \in (a, b))$  también puede escribirse como  $(a < X < b)$  y es una abreviación del evento

$$\{\omega \in \Omega : a < X(\omega) < b\}.$$

Como otro ejemplo considere que el conjunto  $A$  es el intervalo infinito  $(-\infty, x]$  para algún valor real  $x$  fijo. Entonces el evento  $(X \in (-\infty, x])$  también puede escribirse como  $(X \leq x)$  y significa

$$\{\omega \in \Omega : -\infty < X(\omega) \leq x\},$$

que es justamente el conjunto al que se hace referencia en la expresión (2.1). A esta propiedad se le conoce como la condición de **medibilidad** de la función  $X$  respecto de la  $\sigma$ -álgebra  $\mathcal{F}$  del espacio de probabilidad y la  $\sigma$ -álgebra de Borel de  $\mathbb{R}$ . No haremos mayor énfasis en la verificación de esta condición para cada función  $X : \Omega \rightarrow \mathbb{R}$  que se defina pero dicha propiedad es importante pues permite trasladar la medida de probabilidad del espacio de probabilidad a la  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  del siguiente modo:

### Medida de probabilidad inducida

Para cualquier intervalo de la forma  $(-\infty, x]$  se obtiene su imagen inversa bajo  $X$ , es decir,  $X^{-1}(-\infty, x] = \{\omega \in \Omega : X(\omega) \leq x\}$ . Como este conjunto pertenece a  $\mathcal{F}$  por la condición (2.1), se puede aplicar la medida de probabilidad  $P$  pues ésta tiene como dominio  $\mathcal{F}$ . Así, mediante la función  $X$  puede trasladarse la medida de probabilidad  $P$  a intervalos de la forma  $(-\infty, x]$  y puede demostrarse que ello es suficiente para extenderla a la totalidad de la  $\sigma$ -álgebra de Borel de  $\mathbb{R}$ .

A esta nueva medida de probabilidad se le denota por  $P_X(\cdot)$  y se le llama la **medida de probabilidad inducida** por la variable aleatoria  $X$ . Por simplicidad omitiremos el subíndice del término  $P_X$ , de modo que adoptará la misma notación que la medida de probabilidad del espacio de probabilidad original  $(\Omega, \mathcal{F}, P)$ . De esta forma tenemos un nuevo espacio de probabilidad  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ , el cual tomaremos como elemento base de ahora en adelante sin hacer mayor énfasis en ello.

Nuestro interés es el estudio de los distintos eventos de la forma  $(X \in A)$  y sus probabilidades en donde  $X$  es una variable aleatoria y  $A$  es un conjunto de Borel de  $\mathbb{R}$ , por ejemplo un intervalo de la forma  $(a, b)$ .

A menudo se escribe simplemente **v.a.** en lugar del término variable aleatoria. Seguiremos también la notación usual de utilizar la letra mayúscula  $X$  para denotar una variable aleatoria cualquiera, es decir,  $X$  es una función de  $\Omega$  en  $\mathbb{R}$ , mientras que la letra minúscula  $x$  denota un número real y representa un posible valor de la variable aleatoria. En general, las variables aleatorias se denotan usando las últimas letras del alfabeto en mayúsculas:  $U, V, W, X, Y, Z$ , y para un valor cualquiera de ellas se usa la misma letra en minúscula:  $u, v, w, x, y, z$ .

**Ejemplo 2.1** Suponga que un experimento aleatorio consiste en lanzar al aire una moneda equilibrada y observar la cara superior una vez que la moneda cae. Denotemos por “Cara” y “Cruz” los dos lados de la moneda. Entonces claramente el espacio muestral es el conjunto  $\Omega = \{\text{“Cara”}, \text{“Cruz”}\}$ . Defina la variable aleatoria  $X : \Omega \rightarrow \mathbb{R}$  de la siguiente forma:

$$\begin{aligned} X(\text{“Cara”}) &= 0, \\ X(\text{“Cruz”}) &= 1. \end{aligned}$$

De este modo podemos suponer que el experimento aleatorio tiene dos valores numéricos: 0 y 1. Observe que estos números son arbitrarios pues cualquier otro par de números pueden ser escogidos como los valores de  $X$ . Se muestran a continuación algunos ejemplos de eventos de esta variable aleatoria y sus correspondientes probabilidades:

- a)  $P(X \in [1, 2)) = P(\{\text{“Cruz”}\}) = 1/2$ .
- b)  $P(X \in [0, 1)) = P(\{\text{“Cara”}\}) = 1/2$ .
- c)  $P(X \in [2, 4]) = P(\emptyset) = 0$ .
- d)  $P(X = 1) = P(\{\text{“Cruz”}\}) = 1/2$ .
- e)  $P(X \leq -1) = P(\emptyset) = 0$ .
- f)  $P(X \geq 0) = P(\Omega) = 1$ .

•

**Ejemplo 2.2** Sea  $c$  una constante. Para cualquier experimento aleatorio con espacio muestral  $\Omega$  se puede definir la función  $X(\omega) = c$ . Así, cualquier resultado del experimento aleatorio produce, a través de la función  $X$ , el número  $c$ . Decimos entonces que  $X$  es la variable aleatoria constante  $c$  y se puede verificar que para cualquier conjunto de Borel  $A$  de  $\mathbb{R}$ ,

$$P(X \in A) = \begin{cases} 1 & \text{si } c \in A, \\ 0 & \text{si } c \notin A. \end{cases}$$

•

**Ejemplo 2.3** Consideremos el experimento aleatorio consistente en lanzar un dardo en un tablero circular de radio uno como se muestra en la Figura 2.2. El espacio muestral o conjunto de posibles resultados de este experimento se puede escribir como  $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$  y su representación gráfica se muestra en la Figura 2.2.

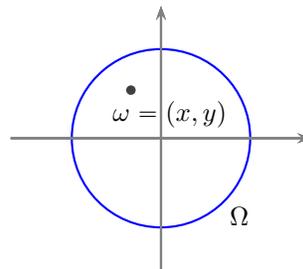


Figura 2.2

Los siguientes son ejemplos de variables aleatorias, es decir, funciones de  $\Omega$  en  $\mathbb{R}$ , asociadas a este experimento aleatorio. Para cualquier  $(x, y) \in \Omega$  se define:

a)  $X(x, y) = x$ .

Esta función es la proyección sobre el eje horizontal. El conjunto de valores que la variable  $X$  puede tomar es el intervalo  $[-1, 1]$ .

b)  $Y(x, y) = y$ .

Esta función es la proyección sobre el eje vertical. El conjunto de valores que la variable  $Y$  puede tomar es el intervalo  $[-1, 1]$ .

c)  $Z(x, y) = \sqrt{x^2 + y^2}$ .

Esta función es la distancia al centro del círculo. El conjunto de valores que la variable  $Z$  puede tomar es el intervalo  $[0, 1]$ .

d)  $V(x, y) = |x| + |y|$ .

Esta función es la así llamada “distancia del taxista”. El conjunto de valores que la variable  $V$  puede tomar es el intervalo  $[0, \sqrt{2}]$ .

e)  $W(x, y) = xy$ .

Esta función es el producto de las coordenadas. El conjunto de valores que la variable  $W$  puede tomar es el intervalo  $[-1/2, 1/2]$ .

•

Otros ejemplos de variables aleatorias aparecen en la sección de ejercicios y muchas otras variables se definirán a lo largo del texto. Nuestro objetivo es el estudio de las distintas variables aleatorias pues éstas codifican en números los resultados de los experimentos aleatorios.

### VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS

Considerando el conjunto de valores que una variable aleatoria puede tomar, vamos a clasificar a las variables aleatorias en dos tipos: discretas o continuas. Decimos que una variable aleatoria es **discreta** cuando el conjunto de valores que ésta toma es un conjunto discreto, es decir, un conjunto finito o numerable. Por ejemplo, el conjunto  $\{0, 1, 2, \dots, n\}$  es un conjunto discreto porque es finito, lo mismo  $\mathbb{N}$  pues aunque éste es un conjunto infinito, es numerable y por lo tanto discreto. Por otra parte, decimos que una variable aleatoria es **continua** cuando toma todos los valores dentro de un intervalo  $(a, b) \subseteq \mathbb{R}$ . Esta clasificación de variables aleatorias no es completa pues existen variables que no son de ninguno de los dos tipos mencionados. Sin embargo, por simplicidad, nos concentraremos en estudiar únicamente variables aleatorias de estos dos tipos.

**Ejemplo 2.4** La variable  $X$  definida en el Ejemplo 2.1 del lanzamiento de una moneda es una variable aleatoria discreta. En ese ejemplo el espacio muestral mismo es discreto y por lo tanto las variables aleatorias que pueden allí definirse tienen que ser discretas forzosamente. En el Ejemplo 2.3 del lanzamiento de un dardo en un tablero circular de radio uno, el espacio muestral de la Figura 2.2 es infinito no numerable, las variables  $X, Y, Z, V$  y  $W$  definidas allí son todas variables aleatorias continuas. Si se dibujan círculos concéntricos alrededor del origen y si se asignan premios asociados a cada una de las regiones resultantes, puede obtenerse un ejemplo de una variable aleatoria discreta sobre este espacio muestral. •

**Ejemplo 2.5** Un experimento aleatorio consiste en escoger a una persona  $\omega$  al azar dentro de una población humana dada. La variable aleatoria  $X$  evaluada en  $\omega$  corresponde a conocer una de las características que aparecen en la lista de abajo acerca de la persona escogida. En cada caso puede considerarse que la variable  $X$  es discreta.

- |                     |              |
|---------------------|--------------|
| a) Edad en años.    | c) Peso.     |
| b) Número de hijos. | d) Estatura. |

•

En las siguientes secciones vamos a explicar la forma de asociar a cada variable aleatoria dos funciones que nos proveen de información acerca de las características de la variable aleatoria. Estas funciones, llamadas función de densidad y función de distribución, nos permiten representar a un mismo tiempo tanto los valores que puede tomar la variable como las probabilidades de los distintos eventos. Definiremos primero la función de probabilidad para una variable aleatoria discreta, después la función de densidad para una v.a. continua, y finalmente definiremos la función de distribución para ambos tipos de variables aleatorias.

## Ejercicios

137. Suponga que un experimento aleatorio consiste en escoger un número al azar dentro del intervalo  $(0, 1)$ . Para cada resultado  $\omega$  del experimento se expresa a este número en su expansión decimal

$$\omega = 0.a_1a_2a_3 \cdots$$

en donde  $a_i \in \{0, 1, \dots, 9\}$ ,  $i = 1, 2, \dots$ . Para cada una de las siguientes variables aleatorias determine si ésta es discreta o continua y en cualquier caso establezca el conjunto de valores que puede tomar:

- a)  $X(\omega) = 1 - \omega$ .
- b)  $Y(\omega) = a_1$ .
- c)  $Z(\omega) = 0.0a_1a_2a_3 \cdots$

138. **Imagen inversa.** Sean  $\Omega_1$  y  $\Omega_2$  dos conjuntos y sea  $X : \Omega_1 \rightarrow \Omega_2$  una función. La imagen inversa de cualquier subconjunto  $A \subseteq \Omega_2$  bajo la función  $X$  es un subconjunto de  $\Omega_1$  denotado por  $X^{-1}A$  y definido como sigue

$$X^{-1}A = \{\omega \in \Omega_1 : X(\omega) \in A\}.$$

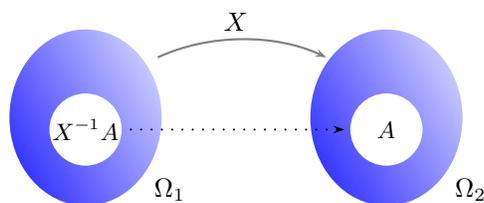


Figura 2.3

Véase la Figura 2.3 y observe que  $X$  es una función puntual mientras que  $X^{-1}$  es una función conjuntista pues lleva subconjuntos de  $\Omega_2$  en subconjuntos de  $\Omega_1$ . Demuestre que la imagen inversa cumple las siguientes propiedades: para  $A, A_1, A_2$  subconjuntos de  $\Omega_2$ ,

- a)  $X^{-1}A^c = (X^{-1}A)^c$ .
- b) Si  $A_1 \subseteq A_2$  entonces  $X^{-1}A_1 \subseteq X^{-1}A_2$ .
- c)  $X^{-1}(A_1 \cup A_2) = (X^{-1}A_1) \cup (X^{-1}A_2)$ .
- d)  $X^{-1}(A_1 \cap A_2) = (X^{-1}A_1) \cap (X^{-1}A_2)$ .

139. **Variable aleatoria constante.** Demuestre que la función  $X : \Omega \rightarrow \mathbb{R}$  que es idénticamente constante  $c$  es una variable aleatoria.

*Sugerencia:* Compruebe que para cualquier número  $x$ ,  $(X \leq x) \in \mathcal{F}$ .

140. **Función indicadora.** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y sea  $A$  un evento. Defina la variable aleatoria  $X$  como aquella que toma el valor 1 si  $A$  ocurre y toma el valor 0 si  $A$  no ocurre. Es decir,

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Demuestre que  $X$  es efectivamente una variable aleatoria. A esta variable se le llama función indicadora del evento  $A$  y se le denota también por  $1_A(\omega)$ .

*Sugerencia:* Compruebe que para cualquier número  $x$ ,  $(X \leq x) \in \mathcal{F}$ .

## 2.2. Función de probabilidad

Consideremos primero el caso discreto. Sea  $X$  una variable aleatoria discreta que toma los valores  $x_0, x_1, \dots$  con probabilidades

$$\begin{aligned} p_0 &= P(X = x_0) \\ p_1 &= P(X = x_1) \\ p_2 &= P(X = x_2) \\ &\vdots \\ &\vdots \end{aligned}$$

Esta lista de valores numéricos y sus probabilidades puede ser finita o infinita, pero numerable. La función de probabilidad de  $X$  se define como aquella función que toma estas probabilidades como valores.

**Definición 2.2** Sea  $X$  una variable aleatoria discreta con valores  $x_0, x_1, \dots$ . La **función de probabilidad** de  $X$ , denotada por  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ , se define como sigue

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_0, x_1, \dots \\ 0 & \text{otro caso.} \end{cases} \quad (2.2)$$

En palabras, la función de probabilidad es simplemente aquella función que indica la probabilidad en los distintos valores que toma la variable aleatoria. Puede escribirse también mediante una tabla de la siguiente forma:

$x$	$x_0$	$x_1$	$x_2$	$\dots$
$f(x)$	$p_0$	$p_1$	$p_2$	$\dots$

Recordemos que es importante poder distinguir entre  $X$  y  $x$ , pues conceptualmente son cosas distintas: la primera es una función y la segunda es un número real. Denotaremos generalmente a una función de probabilidad con la letra  $f$  minúscula. A veces escribiremos  $f_X(x)$  y el subíndice nos ayudará a especificar que tal función es la función de probabilidad de la variable  $X$ . Esta notación será particularmente útil cuando consideremos varias variables aleatorias a la vez. A toda función de la forma (2.2) la llamaremos **función de probabilidad**. Observe que se cumplen las siguientes dos propiedades:

- a)  $f(x) \geq 0$ , para toda  $x \in \mathbb{R}$ .
- b)  $\sum_x f(x) = 1$ .

Recíprocamente, a toda función  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  que sea cero excepto en ciertos puntos  $x_0, x_1, \dots$  en donde la función toma valores positivos se le llamará **función de probabilidad** cuando se cumplen las dos propiedades anteriores y sin que haya de por medio una variable aleatoria. Veamos un par de ejemplos.

**Ejemplo 2.6** Considere la variable aleatoria discreta  $X$  que toma los valores 1, 2 y 3, con probabilidades 0.3, 0.5 y 0.2 respectivamente. Observe que no se especifica ni el experimento aleatorio ni el espacio muestral, únicamente los valores de la variable aleatoria y las probabilidades asociadas. La función de probabilidad de  $X$  es

$$f(x) = \begin{cases} 0.3 & \text{si } x = 1, \\ 0.5 & \text{si } x = 2, \\ 0.2 & \text{si } x = 3, \\ 0 & \text{otro caso.} \end{cases}$$

Esta función se muestra gráficamente en la Figura 2.4(a). Alternativamente podemos también expresar esta función mediante la tabla de la Figura 2.4(b). En esta representación se entiende de manera implícita que  $f(x)$  es cero para cualquier valor de  $x$  distinto de 1, 2 y 3. En particular, no debe ser difícil para el lector comprobar que las siguientes probabilidades son correctas:  $P(X \geq 2) = 0.7$ ,  $P(|X| = 1) = 0.3$  y  $P(X < 1) = 0$ . •

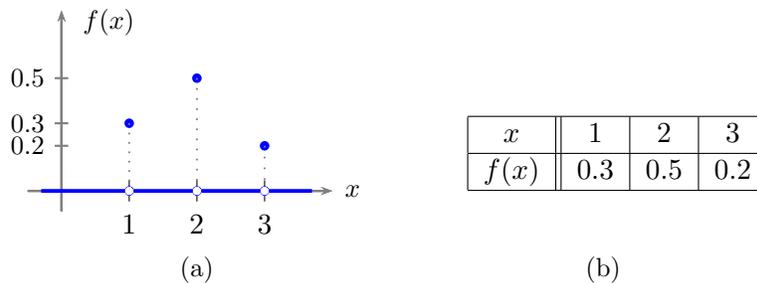


Figura 2.4

**Ejemplo 2.7** Encontraremos el valor de la constante  $c$  que hace que la siguiente función sea de probabilidad.

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

Los posibles valores de la variable aleatoria discreta (no especificada) son 0, 1, 2 y 3, con probabilidades  $0, c, 2c$  y  $3c$ , respectivamente. Como la suma

de estas probabilidades debe ser uno, obtenemos la ecuación  $c + 2c + 3c = 1$ . De aquí obtenemos  $c = 1/6$ . Este es el valor de  $c$  que hace que  $f(x)$  sea no negativa y sume uno, es decir, una función de probabilidad. •

Ahora consideremos el caso continuo. A partir de este momento empezaremos a utilizar el concepto de integral de una función.

**Definición 2.3** Sea  $X$  una variable aleatoria continua. Decimos que la función integrable y no negativa  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  es la **función de densidad** de  $X$  si para cualquier intervalo  $[a, b]$  de  $\mathbb{R}$  se cumple la igualdad

$$P(X \in [a, b]) = \int_a^b f(x) dx.$$

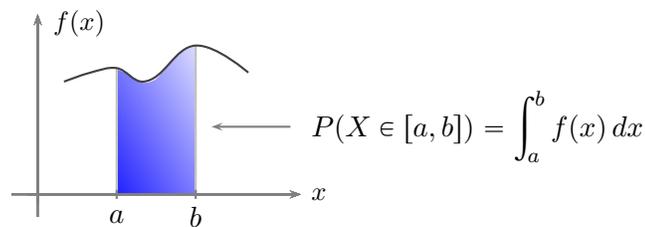


Figura 2.5

Es decir, la probabilidad de que la variable tome un valor dentro del intervalo  $[a, b]$  se puede calcular o expresar como el área bajo la función  $f(x)$  en dicho intervalo. De esta forma el cálculo de una probabilidad se reduce al cálculo de una integral. Véase la Figura 2.5 en donde se muestra esta forma de calcular probabilidades. No es difícil comprobar que toda función de densidad  $f(x)$  de una variable aleatoria continua cumple las siguientes propiedades análogas al caso discreto:

a)  $f(x) \geq 0$ , para toda  $x \in \mathbb{R}$ .

b)  $\int_{-\infty}^{\infty} f(x) dx = 1$ .

Recíprocamente, toda función  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  que satisfaga estas dos propiedades, sin necesidad de tener una variable aleatoria de por medio, se llamará **función de densidad**. Veamos algunos ejemplos.

**Ejemplo 2.8** La función dada por

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

es la función de densidad de una variable aleatoria continua que toma valores en el intervalo  $(1, 3)$  y cuya gráfica aparece en la Figura 2.6. Observe que se trata de una función no negativa y cuya integral vale uno.

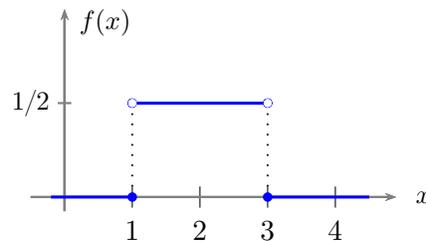


Figura 2.6

**Ejemplo 2.9** Encontraremos el valor de la constante  $c$  que hace que la siguiente función sea de densidad.

$$f(x) = \begin{cases} c|x| & \text{si } x \in [-1, 1], \\ 0 & \text{otro caso.} \end{cases}$$

Se trata de una variable aleatoria continua que toma valores en el intervalo  $[-1, 1]$ . Como esta función debe integrar uno tenemos que

$$1 = \int_{-1}^1 c|x| dx = 2c \int_0^1 x dx = c.$$

Por lo tanto, cuando tomamos  $c = 1$  esta función resulta ser una función de densidad pues ahora cumple con ser no negativa e integrar uno.

## Ejercicios

141. Compruebe que las siguientes funciones son de probabilidad:

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

$$b) f(x) = \begin{cases} \frac{(1-p)p^{x-1}}{1-p^n} & \text{si } x = 1, 2, \dots, n, \quad (0 < p < 1 \text{ constante}) \\ 0 & \text{en otro caso.} \end{cases}$$

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

$$d) f(x) = \begin{cases} \frac{3}{2}(1+x)^2 & \text{si } -1 < x < 0, \\ \frac{3}{2}(1-x)^2 & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

142. En cada caso encuentre el valor de la constante  $\theta$  que hace a la función  $f(x)$  una función de probabilidad. Suponga que  $n$  es un entero positivo fijo.

$$a) f(x) = \begin{cases} \theta x & \text{si } x = 1, 2, \dots, n, \\ 0 & \text{en otro caso.} \end{cases}$$

$$b) f(x) = \begin{cases} \theta x^2 & \text{si } x = 1, 2, \dots, n, \\ 0 & \text{en otro caso.} \end{cases}$$

$$c) f(x) = \begin{cases} \theta 2^x & \text{si } x = 1, 2, \dots, n, \\ 0 & \text{en otro caso.} \end{cases}$$

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

$$e) f(x) = \begin{cases} \theta/3^x & \text{si } x = 1, 3, 5, \dots \\ \theta/4^x & \text{si } x = 2, 4, 6, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

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

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

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

$$i) f(x) = \begin{cases} \theta x + 1/2 & \text{si } -\theta < x < \theta, \\ 0 & \text{en otro caso.} \end{cases}$$

$$j) f(x) = \begin{cases} (1 + \theta)e^{-x} - 2\theta e^{-2x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

143. Sea  $X$  una variable aleatoria discreta con función de probabilidad

$x$	-2	-1	0	1	2
$f(x)$	1/8	1/8	1/2	1/8	1/8

Grafique la función  $f(x)$  y calcule las siguientes probabilidades:

$$a) P(X \leq 1).$$

$$d) P(X^2 \geq 1).$$

$$b) P(|X| \leq 1).$$

$$e) P(|X - 1| < 2).$$

$$c) P(-1 < X \leq 2).$$

$$f) P(X - X^2 - X < 0).$$

144. Sea  $X$  una variable aleatoria con función de densidad

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

Encuentre el valor de la constante  $c$  y calcule:

$$a) P(X \leq 1/2).$$

$$c) P(1/4 < X < 3/4).$$

$$b) P(X \geq 1/4).$$

$$d) a \text{ tal que } P(X \leq a) = 1/2.$$

145. Sea  $X$  una variable aleatoria continua con función de densidad

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

Grafique la función  $f(x)$  y calcule las siguientes probabilidades:

- a)  $P(X \leq 1/3)$ .  
 b)  $P(|X| < 1/2)$ .  
 c)  $P(-1/4 < X < 2/3)$ .  
 d)  $P(2X + 1 < 2)$ .  
 e)  $P(X^2 < 1/9)$ .  
 f)  $P(X^2 - X < 2)$ .

146. Sean  $f(x)$  y  $g(x)$  dos funciones de probabilidad y sea  $c$  una constante. Demuestre o proporcione un contraejemplo para las siguientes afirmaciones:

- a)  $f(x) + g(x)$  es función de probabilidad.  
 b)  $f(x) \cdot g(x)$  es función de probabilidad.  
 c)  $\theta f(x) + (1 - \theta)g(x)$  es función de probabilidad para  $0 \leq \theta \leq 1$ .  
 d)  $f(x + c)$  es función de probabilidad.  
 e)  $f(cx)$  es función de probabilidad.  
 f)  $f(e^x)$  es función de probabilidad.

147. Sea  $X$  una variable aleatoria discreta con valores  $0, 1, 2, \dots$ . Encuentre los valores y la función de probabilidad de la variable:

- a)  $2X$ .  
 b)  $X + n, \quad n \in \mathbb{N}$ .  
 c)  $\cos(2\pi X)$ .  
 d)  $X$  (módulo 2).

148. Un dado equilibrado se lanza dos veces consecutivas. Sea  $X$  la diferencia entre el primer y el segundo lanzamiento. Encuentre la función de probabilidad de  $X$ .

149. Sea  $X$  una variable aleatoria continua con función de densidad

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

- a) Encuentre el valor de la constante  $\theta$  que hace a  $f(x)$  una función de densidad.  
 b) Encuentre los valores de  $a$  y  $b$  tales que  $P(a < X < b) = 1/2$  y  $b - a$  es mínimo.

150. En una caja se encuentran mezcladas 19 baterías en buen estado y 6 baterías defectuosas. Se extraen las baterías al azar una por una hasta encontrar 5 baterías en buen estado. Sea  $X$  la variable aleatoria que indica el número de extracciones efectuadas hasta obtener la quinta batería en buen estado. Encuentre la función de probabilidad de  $X$ .
151. Los jugadores  $A$  y  $B$  llevan a cabo una sucesión de apuestas justas en donde en cada apuesta se gana o se pierde una moneda. Suponga que inicialmente el jugador  $A$  tiene dos monedas y el jugador  $B$  sólo una moneda. El juego termina hasta que uno de los jugadores gana las tres monedas.
- Calcule la probabilidad de ganar de cada uno de los jugadores.
  - Defina la variable aleatoria  $X$  como el número de apuestas efectuadas hasta el final del juego. Calcule la función de probabilidad de  $X$ .
152. Se lanza un dado equilibrado hasta que aparece un número par. Encuentre la función de probabilidad para el número de lanzamientos necesarios hasta obtener el evento de interés.
153. Un experimento consiste en lanzar un dado equilibrado hasta obtener alguno de los resultados por segunda vez. Encuentre la función de probabilidad del número de lanzamientos en este experimento.
154. Una moneda equilibrada marcada con “cara” y “cruz” se lanza repetidamente hasta que se obtienen tres “caras” no necesariamente de manera consecutiva. Encuentre la función de probabilidad de la variable aleatoria que registra el número de lanzamientos necesarios hasta obtener el evento mencionado.
155. Se colocan al azar 10 bolas en 4 cajas de tal forma que cada bola tiene la misma probabilidad de quedar en cualquiera de las 4 cajas. Encuentre la función de probabilidad del número de bolas que caen en la primera caja.
156. **Función de probabilidad condicional** Sea  $X$  una variable aleatoria con función de probabilidad  $f(x)$  y sea  $A \in \mathcal{B}(R)$  un conjunto de Borel

tal que  $p := P(X \in A) > 0$ . La función de probabilidad condicional de  $X$  dado el evento  $(X \in A)$  se denota y define como sigue:

$$\begin{aligned} f(x|A) &:= \frac{1}{p} f(x) \cdot 1_A(x) \\ &= \begin{cases} \frac{1}{p} f(x) & \text{si } x \in A, \\ 0 & \text{en otro caso,} \end{cases} \end{aligned}$$

Demuestre que  $f(x|A)$  es efectivamente una función de probabilidad y calcule esta función para cada uno de los siguientes incisos. En cada caso grafique tanto  $f(x)$  como  $f(x|A)$  y compare ambas funciones.

$$a) \ A = \{2, 3, 4, 5\} \text{ y } f(x) = \begin{cases} 1/6 & \text{si } x = 1, \dots, 6, \\ 0 & \text{en otro caso.} \end{cases}$$

$$b) \ A = (0, \infty) \text{ y } f(x) = \begin{cases} 1/20 & \text{si } -10 < x < 10, \\ 0 & \text{en otro caso.} \end{cases}$$

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

$$d) \ A = (0, 1) \text{ y } f(x) = \begin{cases} x/4 + 1/2 & \text{si } -2 < x \leq 0, \\ -x/4 + 1/2 & \text{si } 0 < x < 2, \\ 0 & \text{en otro caso.} \end{cases}$$

### 2.3. Función de distribución

Otra función importante que puede asociarse a una variable aleatoria es la siguiente.

**Definición 2.4** Sea  $X$  una variable aleatoria cualquiera. La **función de distribución** de  $X$ , denotada por  $F(x) : \mathbb{R} \rightarrow \mathbb{R}$ , se define como la probabilidad

$$F(x) = P(X \leq x). \quad (2.3)$$

Esto es, la función de distribución evaluada en un número  $x$  cualquiera es la probabilidad de que la variable aleatoria tome un valor menor o igual a  $x$ , o en otras palabras, que tome un valor en el intervalo  $(-\infty, x]$ . Para especificar que se trata de la función de distribución de la variable aleatoria  $X$  se usa la notación  $F_X(x)$ , pero por simplicidad omitiremos el subíndice cuando no haya necesidad de tal especificación. Así, siendo  $F(x)$  una probabilidad, sus valores están siempre entre cero y uno. En el caso discreto, suponiendo que  $f(x)$  es la función de probabilidad de  $X$ , la función de distribución (2.3) se calcula como sigue

$$F(x) = \sum_{u \leq x} f(u), \quad (2.4)$$

y corresponde a sumar todos los valores positivos de la función de probabilidad evaluada en aquellos números menores o iguales a  $x$ . En el caso continuo, si  $f(x)$  es la función de densidad de  $X$ , por (2.3) se tiene que

$$F(x) = \int_{-\infty}^x f(u) du. \quad (2.5)$$

La función de distribución resulta ser importante desde el punto de vista matemático pues siempre puede definirse dicha función para cualquier variable aleatoria y a través de ella quedan representadas todas las propiedades de la variable aleatoria. Por razones evidentes se le conoce también con el nombre de **función de acumulación de probabilidad** o **función de probabilidad acumulada**. Mostraremos a continuación algunos ejemplos del cálculo de la función de distribución.

**Ejemplo 2.10 (Caso discreto)** Considere una variable aleatoria discreta  $X$  con función de probabilidad

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

La gráfica de esta función se muestra en la Figura 2.7(a).

Considerando los distintos valores para  $x$ , puede encontrarse la función de distribución  $F(x)$  de la siguiente forma:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) = \begin{cases} 0 & \text{si } x < 1, \\ 1/3 & \text{si } 1 \leq x < 2, \\ 2/3 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3, \end{cases}$$

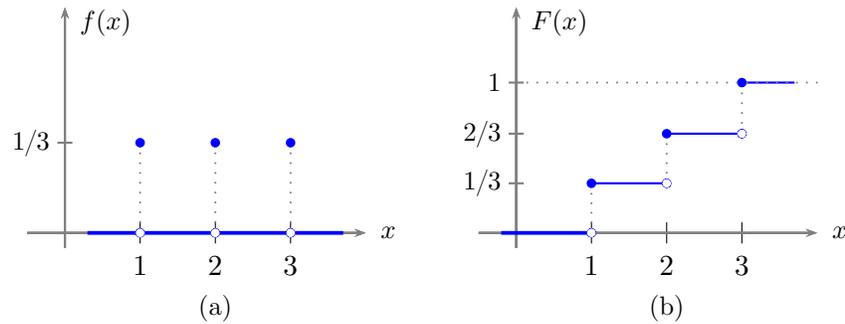


Figura 2.7

cuya gráfica aparece en la Figura 2.7 (b). •

En el ejemplo anterior se ha mostrado el comportamiento típico de una función de distribución discreta, es decir, es una función no decreciente, constante por pedazos, y si la función tiene una discontinuidad en  $x$ , entonces esta discontinuidad es un salto hacia arriba y el tamaño del salto es la probabilidad de que la variable aleatoria tome ese valor.

**Ejemplo 2.11 (Caso continuo)** Considere ahora la variable aleatoria continua  $X$  con función de densidad

$$f(x) = \begin{cases} |x| & \text{si } x \in [-1, 1], \\ 0 & \text{otro caso.} \end{cases}$$

La gráfica de esta función se muestra en la Figura 2.8(a).

Integrando esta función desde menos infinito hasta  $x$ , para distintos valores de  $x$ , se encuentra la función de distribución

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du = \begin{cases} 0 & \text{si } x < -1, \\ (1 - x^2)/2 & \text{si } -1 \leq x < 0, \\ (1 + x^2)/2 & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1, \end{cases}$$

cuya gráfica se muestra en la Figura 2.8(b). •

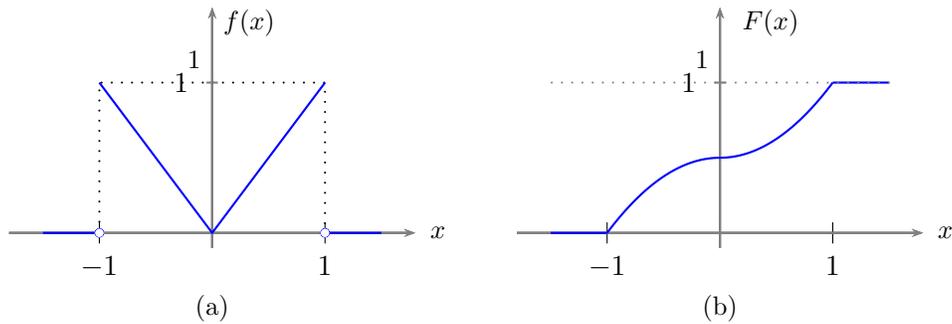


Figura 2.8

En los ejemplos anteriores se ha mostrado la forma de obtener  $F(x)$  a partir de  $f(x)$  tanto en el caso discreto como en el continuo usando las fórmulas (2.4) y (2.5). Ahora explicaremos el proceso contrario, es decir, explicaremos cómo obtener  $f(x)$  a partir de  $F(x)$ . En el caso continuo tenemos que para toda  $x$  en  $\mathbb{R}$ ,

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du,$$

de modo que por el teorema fundamental del cálculo, y cuando  $F(x)$  es diferenciable,

$$f(x) = F'(x). \quad (2.6)$$

De este modo podemos encontrar  $f(x)$  a partir de  $F(x)$ . En el caso discreto, la función de probabilidad se obtiene de la función de distribución del siguiente modo:

$$f(x) = P(X = x) = F(x) - F(x-), \quad (2.7)$$

en donde  $F(x-)$  es el límite por la izquierda de la función  $F$  en el punto  $x$ . Así,  $f(x)$  es efectivamente el tamaño del salto de la función de distribución en el punto  $x$ . Análogamente, la expresión  $F(x+)$  significa el límite por la derecha de la función  $F$  en el punto  $x$ . En símbolos,

$$\begin{aligned} F(x-) &= \lim_{h \searrow 0} F(x-h), \\ F(x+) &= \lim_{h \searrow 0} F(x+h). \end{aligned}$$

Así, haciendo referencia a la ecuación (2.7), cuando la función de distribución es continua, los valores  $F(x)$  y  $F(x-)$  coinciden y la función de probabilidad toma el valor cero. Cuando  $F(x)$  presenta una discontinuidad, la función de probabilidad toma como valor la magnitud de dicha discontinuidad. En los siguientes ejemplos se muestra la aplicación de las fórmulas (2.6) y (2.7).

**Ejemplo 2.12** Considere la función de distribución

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

Se deja al lector la graficación cuidadosa de esta función. Observamos que se trata de una función continua y diferenciable excepto en  $x = -1, 1$ . Derivando entonces en cada una de las tres regiones de definición se encuentra que

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

•

**Ejemplo 2.13** Considere la función de distribución

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

Al graficar esta función uno puede darse cuenta que se trata de una función constante por pedazos. Los puntos donde esta función tiene incrementos y los tamaños de estos incrementos determinan la correspondiente función de probabilidad, la cual está dada por

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

•

Mostraremos a continuación algunas propiedades generales válidas para toda función de distribución. Haremos uso de la propiedad de continuidad de las medidas de probabilidad estudiada en la sección 1.17 del capítulo anterior.

**Proposición 2.1** Toda función de distribución  $F(x)$  satisface las siguientes propiedades:

- a)  $\lim_{x \rightarrow \infty} F(x) = 1$ .
- b)  $\lim_{x \rightarrow -\infty} F(x) = 0$ .
- c) Si  $x_1 \leq x_2$ , entonces  $F(x_1) \leq F(x_2)$ .
- d)  $F(x) = F(x+)$ .

**Demostración.**

- a) Sea  $x_1 \leq x_2 \leq \dots$  cualquier sucesión monótona no decreciente de números reales divergente a infinito. Para cada natural  $n$  defina el evento  $A_n = (X \leq x_n)$ , cuya probabilidad es  $P(A_n) = F(x_n)$ . Entonces  $A_1 \subseteq A_2 \subseteq \dots$  y puede comprobarse que el límite de esta sucesión monótona de eventos es  $\Omega$ . Por la propiedad de continuidad de la probabilidad para sucesiones monótonas,

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

- b) Considere ahora cualquier sucesión monótona no creciente de números reales  $\dots \leq x_2 \leq x_1$  divergente a menos infinito. Defina nuevamente los eventos  $A_n = (X \leq x_n)$  y observe que  $P(A_n) = F(x_n)$ . Entonces  $A_1 \supseteq A_2 \supseteq \dots$  y puede comprobarse que el límite de esta sucesión monótona de eventos es  $\emptyset$ . Por la propiedad de continuidad de la probabilidad para sucesiones monótonas,

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

- c) Si  $x_1 \leq x_2$  entonces  $(X \leq x_1) \subseteq (X \leq x_2)$ . Por lo tanto  $P(X \leq x_1) \subseteq P(X \leq x_2)$ , es decir,  $F(x_1) \leq F(x_2)$ .
- d) Sea  $0 \leq \dots \leq x_2 \leq x_1$  cualquier sucesión monótona no creciente de números reales no negativos convergente a cero. Defina los eventos  $A_n = (x < X \leq x + x_n)$  y observe que  $P(A_n) = F(x + x_n) - F(x)$ . Entonces  $A_1 \supseteq A_2 \supseteq \dots$  y puede comprobarse que el límite de esta sucesión monótona de eventos es  $\emptyset$ . Por la propiedad de continuidad de la probabilidad para sucesiones monótonas,

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

•

Recíprocamente, toda función  $F(x) : \mathbb{R} \rightarrow \mathbb{R}$  que cumpla las cuatro propiedades anteriores (sin tener de por medio una variable aleatoria que la defina) se le llama **función de distribución**. La propiedad (c) significa que  $F(x)$  es una función monótona no decreciente, mientras que la propiedad (d) establece que  $F(x)$  es una función continua por la derecha. Observe que si  $a \leq b$ , entonces se cumplen las siguientes identidades:

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

**Ejemplo 2.14** Sea  $X$  una variable aleatoria con función de distribución

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

Como un ejemplo del cálculo de probabilidades usando la función de distribución, verifique el lector los siguientes resultados:



- a)  $P(X \leq 3)$ .  
 b)  $P(X = 3)$ .  
 c)  $P(X < 3)$ .
- d)  $P(X \geq 1)$ .  
 e)  $P(-1/2 < X < 4)$ .  
 f)  $P(X = 5)$ .

159. Muestre que las siguientes funciones son de probabilidad y encuentre la correspondiente función de distribución. Grafique ambas funciones.

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

$$b) f(x) = \begin{cases} 2x & \text{si } 0 \leq x \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

$$c) f(x) = \begin{cases} 2(1-x) & \text{si } 0 \leq x \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

$$d) f(x) = \begin{cases} 10e^{-10x} & \text{si } x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

160. Encuentre la correspondiente función de distribución para cada una de las funciones de probabilidad que aparecen en la Figura 2.9 en la página 121.

161. Grafique cada una de las siguientes funciones y compruebe que son funciones de distribución. Determine en cada caso si se trata de la función de distribución de una variable aleatoria discreta o continua. Encuentre además la correspondiente función de probabilidad o de densidad.

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

$$b) F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1/5 & \text{si } 0 \leq x < 1, \\ 3/5 & \text{si } 1 \leq x < 2, \\ 1 & \text{si } x \geq 2. \end{cases}$$

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

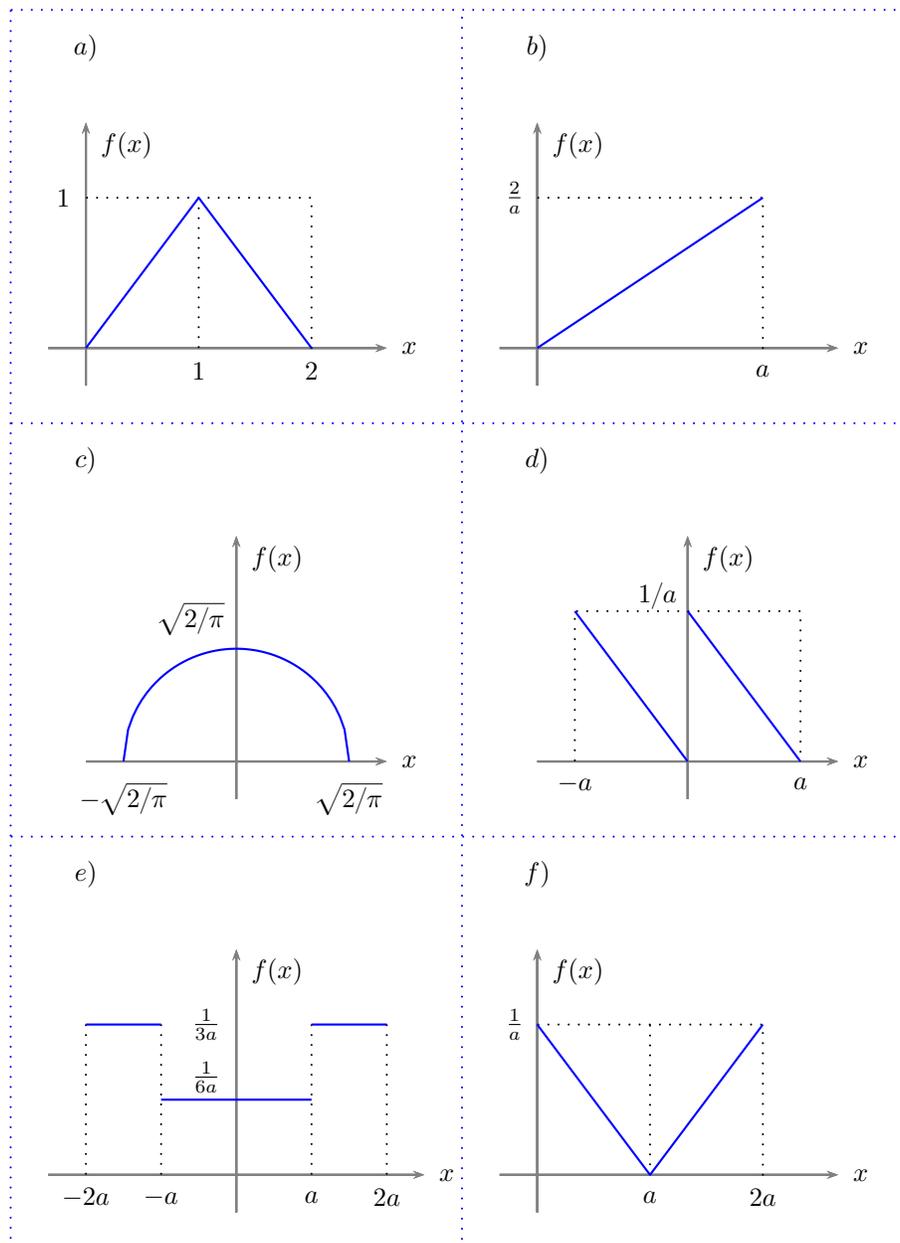


Figura 2.9

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

162. Sea  $X$  una variable aleatoria continua con función de densidad

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

- Grafique  $f(x)$  y compruebe que es efectivamente una función de densidad.
- Calcule y grafique la función de distribución.
- Calcule  $P(|X| < 1/2)$  y  $P(-1/4 < X < 2/3)$ .

163. Sea  $X$  una variable aleatoria con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ x/4 & \text{si } 0 \leq x < 1, \\ 1/2 + (x-1)/4 & \text{si } 1 \leq x < 2, \\ 4/5 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3. \end{cases}$$

- Encuentre  $P(X = x)$  para  $x = 1, 2, 3$ .
- Calcule  $P(1/2 < X < 5/2)$ .

164. Sean  $F(x)$  y  $G(x)$  dos funciones de distribución. Demuestre que para cualquier constante  $\lambda \in (0, 1)$ , la función

$$\lambda F(x) + (1 - \lambda)G(x)$$

es una función de distribución. Si  $F(x)$  y  $G(x)$  son ambas discretas entonces la función resultante es también discreta. Si  $F(x)$  y  $G(x)$  son ambas continuas entonces la función resultante es continua. Si alguna de  $F(x)$  y  $G(x)$  es discreta y la otra es continua, la función resultante no es ni discreta ni continua, se dice que es una función de distribución [mixta](#).

165. Sea  $X$  una variable aleatoria continua con función de distribución

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

Sea  $m > 0$  una constante. Encuentre y grafique la función de distribución de la variable aleatoria:

- a)  $Y = \text{máx}\{X, m\}$ .
- b)  $Y = \text{mín}\{X, m\}$ .

166. Sea  $X$  una variable aleatoria continua con función de densidad  $f_X(x)$  y sea  $c > 0$  una constante. Demuestre que la función de densidad de  $cX$  está dada por

$$f_{cX}(x) = \frac{1}{c} f_X(x/c).$$

167. Sea  $X$  una variable aleatoria continua con función de densidad  $f_X(x)$ . Demuestre que la función de densidad de  $X^2$  está dada por

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}} (f_X(\sqrt{x}) + f_X(-\sqrt{x})) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

168. Demuestre o proporcione un contraejemplo.

- a) Si  $X \leq Y$  entonces  $F_Y(x) \leq F_X(x)$  para toda  $x \in \mathbb{R}$ .
- b) Si  $X = Y$  entonces  $F_X(x) = F_Y(x)$  para toda  $x \in \mathbb{R}$ .
- c) Si  $F_X(x) = F_Y(x)$  para toda  $x \in \mathbb{R}$  entonces  $X = Y$ .

## 2.4. Teorema de cambio de variable

Sea  $X$  una variable aleatoria con distribución conocida. Suponga que se modifica  $X$  a través de una función  $\varphi$  de tal forma que la composición  $\varphi(X)$  es una nueva variable aleatoria, por ejemplo,

$$\begin{aligned} \varphi(X) &= aX + b, \quad a, b \text{ constantes,} \\ \varphi(X) &= X^2, \\ \varphi(X) &= e^X. \end{aligned}$$

El problema natural que surge es el de encontrar la distribución de esta nueva variable aleatoria. En esta sección estudiaremos un resultado que permite dar una respuesta a este problema bajo ciertas condiciones.

**Proposición 2.2** Sea  $X$  una variable aleatoria continua con función de densidad  $f_X(x)$ . Sea  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  una función continua estrictamente creciente o decreciente y con inversa  $\varphi^{-1}$  diferenciable. Entonces la función de densidad de  $Y = \varphi(X)$  está dada por

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \left| \frac{d}{dy} \varphi^{-1}(y) \right| & \text{si } y \in \text{Rango}(\varphi(X)), \\ 0 & \text{en otro caso.} \end{cases} \quad (2.8)$$

**Demostración.** Supondremos el caso cuando  $\varphi$  es estrictamente creciente. El otro caso es análogo. Calcularemos primero la función de distribución de  $Y$  en términos de la función de distribución de  $X$ . Para cualquier valor  $y$  dentro del rango de la función  $Y = \varphi(X)$ ,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\varphi(X) \leq y) \\ &= P(X \leq \varphi^{-1}(y)) \\ &= F_X(\varphi^{-1}(y)). \end{aligned}$$

Derivando respecto de  $y$ , por la regla de la cadena tenemos que

$$f_Y(y) = f_X(\varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y).$$

Como  $\varphi$  es estrictamente creciente, su inversa  $\varphi^{-1}$  también lo es y su derivada es positiva. •

Por simplicidad en el enunciado del resultado anterior hemos pedido que la función  $\varphi$  sea estrictamente monótona y definida de  $\mathbb{R}$  en  $\mathbb{R}$ , sin embargo únicamente se necesita que esté definida en el rango de la función  $X$  y que presente el comportamiento monótono en dicho subconjunto. Ilustraremos esta situación en los siguientes ejemplos.

**Ejemplo 2.15** Sea  $X$  una variable aleatoria continua con función de densidad

$$f_X(x) = \begin{cases} 1 & \text{si } x \in (0, 1), \\ 0 & \text{en otro caso.} \end{cases}$$

Así, la variable  $X$  toma valores únicamente en el intervalo  $(0, 1)$ . Consideremos la función  $\varphi(x) = x^2$ , la cual es estrictamente creciente en  $(0, 1)$ , cuya inversa en dicho intervalo es  $\varphi^{-1}(y) = \sqrt{y}$  y con derivada

$$\frac{d}{dy}\varphi^{-1}(y) = \frac{1}{2\sqrt{y}}.$$

Por lo tanto la variable  $Y = \varphi(X)$  toma valores en  $(0, 1)$  y por la fórmula (2.8) tiene como función de densidad

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & \text{si } y \in (0, 1), \\ 0 & \text{en otro caso.} \end{cases}$$

No es difícil verificar que ésta es efectivamente una función de densidad. •

**Ejemplo 2.16** Sea  $X$  una variable aleatoria continua con función de densidad

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

La variable  $X$  toma valores en el intervalo  $(0, \infty)$ . Sea la función  $\varphi(x) = 1/x$ , la cual es estrictamente decreciente en  $(0, \infty)$ . Su inversa en dicho intervalo es  $\varphi^{-1}(y) = 1/y$  y con derivada

$$\frac{d}{dy}\varphi^{-1}(y) = -\frac{1}{y^2}.$$

Por lo tanto la variable  $Y = \varphi(X)$  toma valores en  $(0, \infty)$  y por la fórmula (2.8) tiene como función de densidad

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

Haciendo el cambio de variable  $u = 1/y$  en la integral puede verificarse con facilidad ésta es efectivamente una función de densidad. •

## Ejercicios

169. Encuentre una fórmula para la función de densidad  $Y$  en términos de la función de densidad de  $X$  cuando esta última es continua con función de densidad  $f_X(x)$  y

a)  $Y = aX + b$  con  $a \neq 0, b$  constantes.

b)  $Y = -X$ .

c)  $Y = X^2$ .

d)  $Y = X^3$ .

e)  $Y = e^X$ .

170. Sea  $X$  una variable aleatoria continua con función de densidad

1)  $f_X(x) = \begin{cases} x^2 + x + 1/6 & \text{si } x \in (0, 1), \\ 0 & \text{en otro caso.} \end{cases}$

2)  $f_X(x) = \begin{cases} 2(1 - |2x - 1|) & \text{si } x \in (0, 1), \\ 0 & \text{en otro caso.} \end{cases}$

En cada caso encuentre la función de densidad de  $Y = \varphi(X)$  cuando

a)  $\varphi(x) = ax + b$  con  $a \neq 0, b$  constantes.

b)  $\varphi(x) = \sqrt{x}$ .

c)  $\varphi(x) = 8(x - 1/2)^3$ .

d)  $\varphi(x) = 1/x$ .

e)  $\varphi(x) = \ln(x)$ .

## 2.5. Independencia de variables aleatorias

El concepto de independencia de eventos puede extenderse al caso de variables aleatorias de una forma natural como la que se presenta en la definición que aparece abajo. En esta sección estudiaremos brevemente este concepto, el cual se revisará con más detalle en la sección 4.6.

**Definición 2.5** Se dice que las variables aleatorias  $X$  y  $Y$  son **independientes** si los eventos  $(X \leq x)$  y  $(Y \leq y)$  son independientes para cualesquiera valores reales de  $x$  y  $y$ , es decir, si para cualquier  $x, y \in \mathbb{R}$  se cumple la igualdad

$$P[(X \leq x) \cap (Y \leq y)] = P(X \leq x) P(Y \leq y). \quad (2.9)$$

El lado izquierdo de la identidad anterior también puede escribirse como  $P(X \leq x, Y \leq y)$  o bien  $F_{X,Y}(x, y)$ , y se le llama la **función de distribución conjunta** de  $X$  y  $Y$  evaluada en el punto  $(x, y)$ . Así, observe que la identidad (2.9) adquiere la expresión

$$F_{X,Y}(x, y) = F_X(x) F_Y(y), \quad x, y \in \mathbb{R}. \quad (2.10)$$

De esta forma, para poder determinar si dos variables aleatorias son independientes es necesario conocer tanto las probabilidades conjuntas  $P(X \leq x, Y \leq y)$  como las probabilidades individuales  $P(X \leq x)$  y  $P(Y \leq y)$ , y verificar la identidad (2.10) para cada par de números reales  $x$  y  $y$ . Por lo tanto basta que exista una pareja  $(x, y)$  para la que no se cumpla la igualdad (2.10) para poder concluir que  $X$  y  $Y$  no son independientes. En el capítulo de vectores aleatorios explicaremos la forma de obtener las distribuciones individuales a partir de la distribución conjunta. Nuestro objetivo por ahora es mencionar que no es difícil demostrar, y en el ejercicio 171 se pide hacerlo, que cuando  $X$  y  $Y$  son discretas, la condición (2.10) es equivalente a la siguiente expresión simple:

**Independencia de variables aleatorias: caso discreto**

$$P(X = x, Y = y) = P(X = x) P(Y = y), \quad x, y \in \mathbb{R},$$

en donde, por el teorema de probabilidad total,

$$P(X = x) = \sum_y P(X = x, Y = y),$$

$$P(Y = y) = \sum_x P(X = x, Y = y).$$

En el caso cuando  $X$  y  $Y$  son continuas y la función  $F_{X,Y}(x, y)$  puede expresarse de la siguiente forma

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

entonces a  $f_{X,Y}(x, y)$  se le llama la **función de densidad conjunta** de  $X$  y  $Y$ , y puede demostrarse que la condición de independencia (2.10) es equivalente a la siguiente expresión:

**Independencia de variables aleatorias: caso continuo**

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad x, y \in \mathbb{R},$$

en donde,

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

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

La definición de independencia de dos v.a.s puede extenderse de manera análoga al caso de tres o más variables aleatorias. Una discusión más completa de este tema puede encontrarse en el capítulo sobre vectores aleatorios.

**Ejemplo 2.17 (Producto Cartesiano de espacios muestrales)** El siguiente procedimiento es una versión simple de un mecanismo general para construir variables aleatorias independientes: consideremos que tenemos un primer experimento aleatorio con espacio muestral  $\Omega_1 = \{a_0, a_1\}$  y con probabilidades  $p$  y  $1 - p$  para cada uno de estos resultados respectivamente, y que por otro lado también tenemos un segundo experimento con espacio muestral  $\Omega_2 = \{b_0, b_1\}$  con probabilidades  $q$  y  $1 - q$  para estos dos resultados.

La intención es crear el experimento aleatorio consistente en llevar a cabo el primer experimento seguido del segundo y para este experimento compuesto definir dos variables aleatorias independientes. Así, para el experimento compuesto el espacio muestral será el espacio producto:

$$\Omega = \Omega_1 \times \Omega_2 = \{(a_0, b_0), (a_0, b_1), (a_1, b_0), (a_1, b_1)\},$$

y definamos la probabilidad de cada uno de estos cuatro resultados como el producto de las probabilidades individuales, es decir, se asocian las probabilidades:  $pq, p(1-q), (1-p)q, (1-p)(1-q)$ , respectivamente conforme al orden en el que aparecen en el conjunto muestral. Esta asignación de probabilidades es el elemento clave para verificar la propiedad de independencia. Defina ahora las variables aleatorias  $X$  y  $Y$  de la siguiente forma:

$$\begin{aligned} X(a_0, -) &= 0, & Y(-, b_0) &= 0, \\ X(a_1, -) &= 1, & Y(-, b_1) &= 1. \end{aligned}$$

Por lo tanto, sin importar el resultado del segundo experimento la variable  $X$  asigna los valores cero y uno a los dos resultados del primer experimento. Simétricamente, sin importar el resultado del primer experimento la variable  $Y$  asigna los valores cero y uno a los dos resultados del segundo experimento. Puede verificarse que  $X$  y  $Y$  son independientes pues por construcción para cualesquiera valores de  $x$  y  $y$  se cumple la identidad

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

•

## Ejercicios

171. **Independencia: caso discreto.** Sean  $X$  y  $Y$  dos variables aleatorias discretas. Sin pérdida de generalidad suponga que ambas toman valores en subconjuntos de  $\mathbb{Z}$ . Compruebe las siguientes identidades:

$$a) \quad P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} P(X = u, Y = v).$$

$$\begin{aligned}
 b) \quad P(X = x, Y = y) &= P(X \leq x, Y \leq y) \\
 &\quad - P(X \leq x - 1, Y \leq y) \\
 &\quad - P(X \leq x, Y \leq y - 1) \\
 &\quad + P(X \leq x - 1, Y \leq y - 1).
 \end{aligned}$$

Con estos resultados demuestre que la condición de independencia

$$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y), \quad x, y \in \mathbb{R}.$$

es equivalente a la condición

$$P(X = x, Y = y) = P(X = x) P(Y = y), \quad x, y \in \mathbb{R}.$$

172. Sean  $X_1, \dots, X_n$  variables aleatorias discretas e independientes con idéntica función de probabilidad dada por

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

Encuentre la función de probabilidad de la variable aleatoria:

$$a) \quad U = \text{máx}\{X_1, \dots, X_n\}.$$

$$b) \quad V = \text{mín}\{X_1, \dots, X_n\}.$$

## 2.6. Esperanza

En las siguientes secciones estudiaremos ciertas cantidades numéricas que pueden ser calculadas para cada variable aleatoria. Estos números revelan algunas características de la variable aleatoria o de su distribución. La primera de estas características numéricas es la esperanza, se denota por  $E(X)$  y se calcula como se define a continuación.

**Definición 2.6** Sea  $X$  una v.a. discreta con función de probabilidad  $f(x)$ . La **esperanza** de  $X$  se define como

$$E(X) = \sum_x x f(x), \quad (2.11)$$

suponiendo que esta suma es absolutamente convergente, es decir, cuando la suma de los valores absolutos es convergente. Por otro lado, si  $X$  es continua con función de densidad  $f(x)$ , entonces la **esperanza** es

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx, \quad (2.12)$$

suponiendo que esta integral es absolutamente convergente, es decir, cuando la integral de los valores absolutos es convergente.

El número de sumandos en la expresión (2.11) puede ser finito o infinito, dependiendo del conjunto de valores de la variable aleatoria. Además, si la suma o integral anteriores no cumplen esta condición de convergencia absoluta, entonces se dice que la esperanza no existe o bien que no tiene esperanza finita. En los Ejercicios 179, 180 y 181 pueden encontrarse ejemplos en donde se presenta esta situación.

La esperanza de una variable aleatoria es entonces un número que indica el promedio ponderado de los diferentes valores que la variable puede tomar. A la esperanza se le conoce también con los nombres de **media**, **valor esperado** o **valor promedio**. En general se usa también la letra griega  $\mu$  (mu) para denotarla, es uno de los conceptos más importantes en probabilidad y tiene un amplio uso en las aplicaciones y otras ramas de la ciencia. A partir de este momento resultará muy útil conocer algunas fórmulas para llevar a cabo sumas y poseer un manejo adecuado de las técnicas de integración. Mediante algunos ejemplos ilustraremos a continuación la forma de calcular esperanzas de variables aleatorias.

**Ejemplo 2.18 (Caso discreto)** Sea  $X$  una variable aleatoria discreta con función de probabilidad dada por la siguiente tabla:

$x$	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

La esperanza de  $X$  es el número

$$\begin{aligned} E(X) &= \sum_x x f(x) \\ &= (-1)(1/8) + (0)(4/8) + (1)(1/8) + (2)(2/8) \\ &= 1/2. \end{aligned}$$

Observe que la suma se efectúa sobre todos los valores de  $x$  indicados en la tabla, es decir:  $-1, 0, 1$  y  $2$ . También es instructivo observar que la esperanza no es necesariamente uno de los valores tomados por la variable aleatoria. En este ejemplo el valor  $1/2$  nunca es tomado por la variable aleatoria pero es su valor promedio. •

**Ejemplo 2.19 (Caso continuo)** Considere la variable aleatoria continua  $X$  con función de densidad

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

La esperanza de  $X$  es

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}.$$

Observe que la integral sólo es relevante en el intervalo  $(0, 1)$ , pues fuera de este intervalo la función de densidad se anula. •

De este modo la esperanza de una variable aleatoria es un **promedio ponderado** de los valores que toma la variable aleatoria: cada valor  $x$  se pondera con la función de probabilidad  $f(x)$  y se suman todas estas cantidades.

### Esperanza de una función de una variable aleatoria

En algunos casos es necesario calcular la esperanza de una función de una variable aleatoria, por ejemplo, si  $X$  es una variable aleatoria discreta, entonces es claro que  $Y = X^2$  es una función de  $X$  y consideraremos que  $Y$  es también una variable aleatoria. Si quisiéramos encontrar la esperanza de  $Y$  según la expresión (2.11) de la definición de esperanza tendríamos que calcular

$$E(Y) = \sum_y y f_Y(y).$$

Para lo cual se necesita encontrar primero la función de densidad de  $Y$  y ello en general no es fácil. El siguiente resultado es muy útil y nos dice la forma de calcular esta esperanza conociendo únicamente la función de densidad de  $X$ . A este resultado a veces se le refiere como el teorema del estadístico inconsciente.

**Proposición 2.3** Sea  $X$  una variable aleatoria discreta y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  una función tal que  $g(X)$  es una variable con esperanza finita. Entonces

$$E[g(X)] = \sum_x g(x) f_X(x). \quad (2.13)$$

**Demostración.** Sea  $Y = g(X)$  y sea  $y$  uno de sus posibles valores. Así, existe por lo menos un valor  $x$  tal que  $y = g(x)$ . Agrupemos todos estos valores  $x$  en el conjunto  $g^{-1}y = \{x : g(x) = y\}$ . De este modo tenemos que

$$P(Y = y) = P(X \in g^{-1}y) = \sum_{x \in g^{-1}y} f_X(x).$$

Es claro que la unión de los conjuntos disjuntos  $g^{-1}y$ , considerando todos los valores  $y$ , es el conjunto de valores  $x$  que la v.a.  $X$  puede tomar. Por lo

tanto la esperanza de  $Y$  es

$$\begin{aligned}
 E(Y) &= \sum_y yP(Y = y) \\
 &= \sum_y y \sum_{x \in g^{-1}y} f_X(x) \\
 &= \sum_y \sum_{x \in g^{-1}y} g(x)f_X(x) \\
 &= \sum_x g(x)f_X(x).
 \end{aligned}$$

•

Observe que en la fórmula (2.13) no aparece la función de probabilidad de  $Y$  sino la de  $X$ . Allí radica la utilidad de esta fórmula pues para calcular  $E(Y)$  no es necesario conocer  $f_Y(y)$ . Veamos un ejemplo de aplicación.

**Ejemplo 2.20** Sea  $X$  una variable aleatoria discreta con función de probabilidad dada por la siguiente tabla:

$x$	-2	-1	0	1	2
$f_X(x)$	2/8	1/8	2/8	1/8	2/8

Nos interesa calcular la esperanza de la variable aleatoria  $Y = X^2$ .

*Solución 1.* Un primer método consiste en calcular primero la función de probabilidad de la variable  $Y$ . No es complicado verificar que esta función de probabilidad es

$y$	0	1	4
$f_Y(y)$	2/8	2/8	4/8

Entonces usando la definición elemental de esperanza para variables aleatorias discretas tenemos que

$$E(Y) = (0)\left(\frac{2}{8}\right) + (1)\left(\frac{2}{8}\right) + (4)\left(\frac{4}{8}\right) = \frac{9}{4}.$$

*Solución 2.* Ahora calcularemos la misma esperanza pero usando la fórmula (2.13). El resultado es el mismo pero la ventaja es que no es necesario calcular la función de probabilidad de  $Y$ :

$$E(Y) = (-2)^2\left(\frac{2}{8}\right) + (-1)^2\left(\frac{1}{8}\right) + (0)^2\left(\frac{2}{8}\right) + (1)^2\left(\frac{1}{8}\right) + (2)^2\left(\frac{2}{8}\right) = \frac{9}{4}.$$

•

El siguiente resultado es la versión continua de la Proposición 2.3 y su demostración requiere conceptos más avanzados de probabilidad, así es que omitiremos la demostración y nos concentraremos en su uso y aplicación.

**Proposición 2.4** Sea  $X$  una variable aleatoria continua y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  una función tal que  $g(X)$  es una variable con esperanza finita. Entonces

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (2.14)$$

**Ejemplo 2.21** Calcularemos  $E(X^2)$  en donde  $X$  es la variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 1 & \text{para } x \in (0, 1), \\ 0 & \text{en otro caso.} \end{cases}$$

*Solución 1.* Si se desea aplicar la identidad (2.12) de la definición elemental de esperanza, se tiene que encontrar primero la distribución de la v.a.  $X^2$ . Puede verificarse que

$$F_{X^2}(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \sqrt{x} & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

De donde, derivando,

$$f_{X^2}(x) = \begin{cases} \frac{1}{2}x^{-1/2} & \text{si } 0 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

Finalmente aplicando la definición elemental de esperanza (2.12),

$$E(X^2) = \int_0^1 x f_{X^2}(x) dx = \int_0^1 \frac{1}{2} x^{1/2} dx = \frac{1}{3}.$$

*Solución 2.* Ahora resolveremos el mismo problema de una manera más rápida aplicando el resultado de la proposición anterior. Por la fórmula 2.14 tenemos que

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 dx = \frac{1}{3}.$$

•

Estudiaremos a continuación algunas propiedades generales de la esperanza.

**Proposición 2.5 (Propiedades de la esperanza)** Sean  $X$  y  $Y$  dos variables aleatorias con esperanza finita y sea  $c$  una constante. Entonces

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

**Demostración.** La primera propiedad es evidente pues si  $X$  es la variable aleatoria constante  $c$ , entonces por definición,  $E(X) = cP(X = c) = c \cdot 1 = c$ . El segundo inciso se sigue directamente de la definición de esperanza pues tanto en el caso discreto como en el caso continuo, la constante  $c$  puede siempre colocarse fuera de la suma o integral. El tercer inciso también es inmediato pues en la integral o suma correspondiente solo aparecerán términos que son no negativos. Para la última propiedad, consideraremos el caso en el que ambas variables son discretas y por simplicidad usaremos la expresión  $p(x, y)$  para denotar  $P(X = x, Y = y)$ .

$$\begin{aligned}
E(X + Y) &= \sum_{x,y} (x + y) p(x, y) \\
&= \sum_{x,y} x p(x, y) + \sum_{x,y} y p(x, y) \\
&= \sum_x x \sum_y p(x, y) + \sum_y y \sum_x p(x, y) \\
&= \sum_x x p(x) + \sum_y y p(y) \\
&= E(X) + E(Y).
\end{aligned}$$

•

Observe que la segunda y cuarta propiedad establecen que la esperanza es **lineal**, es decir, separa sumas y separa multiplicaciones por constantes. Veamos ahora una aplicación del concepto de independencia en el cálculo de la esperanza.

**Proposición 2.6** Sean  $X$  y  $Y$  independientes y ambas con esperanza finita. Entonces

$$E(XY) = E(X) E(Y).$$

**Demostración.** Por simplicidad consideraremos únicamente el caso en el que ambas variables aleatorias son discretas. En el caso continuo el procedimiento es análogo.

$$\begin{aligned}
E(XY) &= \sum_{x,y} x y P(X = x, Y = y) \\
&= \sum_x \sum_y x y P(X = x) P(Y = y) \\
&= \left( \sum_x x P(X = x) \right) \left( \sum_y y P(Y = y) \right) \\
&= E(X) E(Y).
\end{aligned}$$

•

El recíproco de la proposición anterior no es válido en general, es decir, la condición  $E(XY) = E(X)E(Y)$  no es suficiente para concluir que  $X$  y  $Y$  son independientes. Véase el Ejercicio 189 para un ejemplo de esta situación.

## Ejercicios

173. Calcule la esperanza de la variable aleatoria discreta  $X$  con función de probabilidad:

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

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

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

$$\text{Respuestas: } \begin{cases} a) (n+1)/2. \\ b) 1. \\ c) (2n+1)/3. \end{cases}$$

174. Calcule la esperanza de la variable aleatoria continua  $X$  con función de densidad:

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

$$b) f(x) = \begin{cases} x^2 + \frac{4}{3}x & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

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

$$\text{Respuestas: } \begin{cases} a) 0. \\ b) 25/36. \\ c) 1. \end{cases}$$

175. **Monotonía.** Sean  $X$  y  $Y$  dos variables aleatorias con esperanza finita. Demuestre que si  $X \leq Y$  entonces  $E(X) \leq E(Y)$ .
176. **Fórmula alternativa, caso discreto.** Sea  $X$  una variable aleatoria discreta con función de distribución  $F(x)$ , con esperanza finita y posibles valores dentro del conjunto  $\{0, 1, 2, \dots\}$ . Demuestre que

$$E(X) = \sum_{x=0}^{\infty} (1 - F(x)). \quad (2.15)$$

Use la fórmula anterior para encontrar  $E(X)$  cuando:

$$a) F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1/5 & \text{si } 0 \leq x < 1, \\ 3/5 & \text{si } 1 \leq x < 2, \\ 4/5 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3. \end{cases}$$

$$b) F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1 - (1/2)^k & \text{si } k \leq x < k + 1, \quad k \in \mathbb{N}. \end{cases}$$

177. **Fórmula alternativa, caso continuo.** Sea  $X$  una variable aleatoria continua con función de distribución  $F(x)$ , con esperanza finita y con valores en el intervalo  $[0, \infty)$ . Demuestre que

$$E(X) = \int_0^{\infty} P(X > x) dx. \quad (2.16)$$

Use la fórmula anterior para encontrar  $E(X)$  cuando:

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

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

178. Demuestre o proporcione un contraejemplo:

$$a) E(E(X)) = E^2(X).$$

$$b) E(X^2 - Y^2) = E(X + Y) E(X - Y).$$

- c)  $E(1/X) = 1/E(X)$ .
- d)  $E(X - E(X)) = E(E(X) - X) = 0$ .
- e) Si  $E(X) = 0$  entonces  $X = 0$ .
- f) Si  $E(X^2) = 0$  entonces  $X = 0$ .
- g) Si  $E(X) = E(Y)$  entonces  $X = Y$ .
- h)  $E(X) \leq E(X^2)$ .

179. **Sin esperanza, caso discreto.** Sea  $X$  una variable aleatoria discreta con función de probabilidad  $f(x)$  como aparece abajo. Demuestre que  $f(x)$  es efectivamente una función de probabilidad y compruebe que  $X$  no tiene esperanza finita.

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

180. **Sin esperanza, caso continuo.** Sea  $X$  una variable aleatoria continua con función de densidad  $f(x)$  como aparece abajo. Demuestre que  $f(x)$  es efectivamente una función de densidad y compruebe que  $X$  no tiene esperanza finita.

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

181. **La paradoja de San Petersburgo.** Un jugador lanza repetidas veces una moneda equilibrada hasta obtener una de las caras previamente escogida. El jugador obtiene un premio de  $2^n$  unidades monetarias si logra su objetivo en el  $n$ -ésimo lanzamiento. Calcule el valor promedio del premio en este juego.

Respuesta: Infinito.

182. **El problema del ladrón de Bagdad.** El ladrón de Bagdad ha sido colocado en una prisión en donde hay tres puertas. Una de las puertas conduce a un túnel que requiere de un día de travesía y que lo regresa a la misma prisión. Otra de las puertas lo conduce a otro túnel aún más largo que lo regresa nuevamente a la prisión pero después de tres días de recorrido. Finalmente la tercera puerta lo conduce a la libertad inmediatamente. Suponga que el ladrón escoge al azar cada una

estas puertas una por una hasta quedar libre. Encuentre el número promedio de días que le toma al ladrón quedar en libertad.

Respuesta: 2.

183. Un jugador lanza dos veces consecutivas un dado equilibrado y obtiene como premio tantas unidades monetarias como indica el resultado mayor de los dos lanzamientos. Encuentre el valor promedio del premio.

184. Sea  $X$  una variable aleatoria positiva y con esperanza finita. Demuestre que

$$\frac{1}{E(X)} \leq E\left(\frac{1}{X}\right).$$

185. Sea  $X$  una variable aleatoria con función de densidad  $f(x)$  como aparece abajo. Encuentre el valor de la constante  $a > 0$  tal que  $E(X) = 0$ .

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

186. **Función signo.** La función signo se define de la forma siguiente:

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

Calcule la esperanza de la variable aleatoria  $Y = \text{signo}(X)$  cuando  $X$  tiene función de probabilidad:

a) 

$x$	-2	-1	0	1	2
$f(x)$	1/8	2/8	1/8	2/8	2/8

b)  $f(x) = \begin{cases} (x+1)/2 & \text{si } x \in (-1, 1), \\ 0 & \text{en otro caso.} \end{cases}$

c)  $f(x) = \begin{cases} e^{-(x+a)} & \text{si } x > -a, \\ 0 & \text{en otro caso.} \end{cases} \quad (a > 0 \text{ constante})$

d)  $f(x) = \frac{1}{2}e^{-|x|}$ .

187. Sea  $X$  una variable aleatoria con función de probabilidad  $f(x)$  como aparece abajo. Calcule la esperanza de la variable  $Y = e^X$  encontrando primero la función de densidad de  $Y$  y aplicando la definición elemental de esperanza. Como segundo método use el teorema del estadístico inconsciente. Ambos cálculos deben coincidir.

$$a) f(x) = \begin{cases} (1-p)p^{x-1} & \text{si } x = 1, 2, \dots \quad (0 < p < e^{-1} \text{ constante}) \\ 0 & \text{en otro caso.} \end{cases}$$

$$b) f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0, \quad (\lambda > 1 \text{ constante}) \\ 0 & \text{en otro caso.} \end{cases}$$

$$\text{Respuestas: } \begin{cases} a) E(Y) = \frac{e(1-p)}{1-ep}. \\ b) E(Y) = \frac{\lambda}{\lambda-1}. \end{cases}$$

188. En una población pequeña ocurren cada día 0, 1, 2, o 3 accidentes automovilísticos con probabilidades  $1/3$ ,  $1/3$ ,  $1/6$  y  $1/6$  respectivamente. En un accidente cualquiera se requiere el uso de una ambulancia con probabilidad 0.5. Calcule el número de veces promedio que se requiere el uso de una ambulancia por accidentes automovilísticos en un día cualquiera en esta población.
189. Sea  $X$  una variable aleatoria discreta con función de probabilidad dada por la tabla que aparece abajo. Defina por otro lado a la variable  $Y = X^2$ . Se cumple entonces la condición  $E(XY) = E(X)E(Y)$  y sin embargo  $X$  y  $Y$  no son independientes. Demuestre ambas afirmaciones.

$x$	-1	0	1
$f(x)$	$1/3$	$1/3$	$1/3$

## 2.7. Varianza

Otra característica numérica importante asociada a las variables aleatorias se llama varianza, se denota por  $\text{Var}(X)$  y se calcula de la forma siguiente.

**Definición 2.7** Sea  $X$  una v.a. discreta con función de probabilidad  $f(x)$ . La **varianza** de  $X$  se define como sigue

$$\text{Var}(X) = \sum_x (x - \mu)^2 f(x),$$

cuando esta suma es convergente y en donde  $\mu$  es la esperanza de  $X$ . Para una variable aleatoria continua  $X$  con función de densidad  $f(x)$  se define

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx,$$

cuando esta integral es convergente.

Así, observe que se necesita conocer la esperanza de  $X$  para calcular su varianza. Es interesante observar también que la varianza puede escribirse en una sola expresión como sigue

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

Esto corresponde a la esperanza de la función cuadrática  $x \mapsto (x - \mu)^2$  aplicada a una variable aleatoria  $X$  con esperanza  $\mu$ . La varianza es una medida del grado de **dispersión** de los diferentes valores tomados por la variable. Se le denota regularmente por la letra  $\sigma^2$  (sigma cuadrada). A la raíz cuadrada positiva de la varianza, esto es  $\sigma$ , se le llama **desviación estándar**. Como en el caso de la esperanza, la anterior suma o integral puede no ser convergente y en ese caso se dice que la variable aleatoria no tiene varianza finita. Veamos algunos ejemplos.

**Ejemplo 2.22 (Caso discreto)** Calcularemos la varianza de la variable aleatoria discreta  $X$  con función de probabilidad dada por la siguiente tabla.

$x$	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

Recordemos primeramente que por cálculos previos,  $\mu = 1/2$ . Aplicando la

definición de varianza tenemos que

$$\begin{aligned}\text{Var}(X) &= \sum_x (x - \mu)^2 f(x) \\ &= (-1 - 1/2)^2(1/8) + (0 - 1/2)^2(4/8) \\ &\quad + (1 - 1/2)^2(1/8) + (2 - 1/2)^2(2/8) \\ &= 1.\end{aligned}$$

•

**Ejemplo 2.23 (Caso continuo)** Calcularemos la varianza de la variable aleatoria continua  $X$  con función de densidad

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

En un cálculo previo habíamos encontrado que  $\mu = 2/3$ . Por lo tanto,

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_0^1 (x - 2/3)^2 2x dx = \frac{1}{18}.$$

•

A continuación demostraremos algunas propiedades generales de la varianza.

**Proposición 2.7 (Propiedades de la varianza)** Sean  $X$  y  $Y$  dos variables aleatorias con varianza finita y sea  $c$  una constante. Entonces

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

**Demostración.** El inciso (1) es evidente a partir de la definición de varianza pues en ella aparece una suma o integral de términos no negativos. Para el inciso (2) la constante  $c$  es una v.a. con un único valor, de modo que  $E(c) = c$  y entonces  $\text{Var}(X) = E(c - c)^2 = 0$ . Para el inciso (3) tenemos que

$$\begin{aligned}\text{Var}(cX) &= E(cX - E(cX))^2 \\ &= E(cX - cE(X))^2 \\ &= c^2 E(X - E(X))^2 \\ &= c^2 \text{Var}(X).\end{aligned}$$

El inciso (4) se sigue del siguiente análisis:

$$\text{Var}(X + c) = E[(X + c) - E(X + c)]^2 = E(X - E(X))^2 = \text{Var}(X).$$

Para demostrar la propiedad (5) se desarrolla el cuadrado en la definición de varianza y se usa la propiedad de linealidad de la esperanza:

$$\begin{aligned}\text{Var}(X) &= E(X - E(X))^2 \\ &= E(X^2 - 2XE(X) + E^2(X)) \\ &= E(X^2) - 2E(X)E(X) + E^2(X) \\ &= E(X^2) - E^2(X).\end{aligned}$$

Finalmente para demostrar la propiedad (6) es suficiente dar un ejemplo. Puede tomarse el caso  $Y = X$ , en general y por lo demostrado antes, no se cumple que  $\text{Var}(2X) = 2 \text{Var}(X)$ . •

De estas propiedades generales se obtiene en particular que la varianza es siempre una cantidad no negativa y que no cumple la propiedad de linealidad, pues en general no separa sumas y cuando aparecen constantes como factores, las constantes se separan de la varianza elevándolas al cuadrado. Otras propiedades se encuentran en el Ejercicio 195. Veamos ahora una fórmula para el cálculo de la varianza de la suma de dos variables aleatorias bajo la hipótesis de independencia.

**Proposición 2.8** Sean  $X$  y  $Y$  independientes y ambas con varianza finita. Entonces

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

**Demostración.** Usaremos la propiedad de linealidad de la esperanza y el hecho de que  $E(XY) = E(X)E(Y)$  pues  $X$  y  $Y$  son independientes por hipótesis.

$$\begin{aligned} \text{Var}(X + Y) &= E(X + Y)^2 - E^2(X + Y) \\ &= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) \\ &\quad - E^2(X) - 2E(X)E(Y) - E^2(Y) \\ &= (E(X^2) - E^2(X)) + (E(Y^2) - E^2(Y)) \\ &= \text{Var}(X) + \text{Var}(Y). \end{aligned}$$

•

El recíproco del resultado anterior es en general falso, es decir, la condición  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$  no es suficiente para concluir que  $X$  y  $Y$  son independientes. Un ejemplo de esta situación se muestra en el Ejercicio 377, el cual requiere del concepto de distribución conjunta de variables aleatorias que estudiaremos con más detalle en el capítulo sobre vectores aleatorios.

## Ejercicios

190. Calcule la media y la varianza de la variable aleatoria discreta  $X$  con función de probabilidad:

$$\begin{aligned} a) \quad f(x) &= \begin{cases} 1/n & \text{si } x = 1, 2, \dots, n, \\ 0 & \text{en otro caso.} \end{cases} \\ b) \quad f(x) &= \begin{cases} 1/2^x & \text{si } x = 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

191. Calcule la media y la varianza de la variable aleatoria continua  $X$  con función de densidad:

$$a) f(x) = \begin{cases} 1/3 & \text{si } 0 < |x| < 1, \\ 1/6 & \text{si } 1 < |x| < 2, \\ 0 & \text{en otro caso.} \end{cases}$$

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

$$c) f(x) = \begin{cases} |x| & \text{si } |x| < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

$$a) E(X) = 0 \quad \text{Var}(X) = 5/18.$$

$$\text{Respuestas: } b) E(X) = 1 \quad \text{Var}(X) = 1.$$

$$c) E(X) = 0 \quad \text{Var}(X) = 1/2.$$

192. Encuentre la distribución de la variable aleatoria  $X$  que cumple las siguientes dos condiciones:

$$P(X = 1) = 1 - P(X = -1),$$

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

$$\text{Respuesta: } \begin{aligned} P(X = 1) &= \frac{1+\sqrt{5}}{4}, \\ P(X = -1) &= \frac{3-\sqrt{5}}{4}. \end{aligned}$$

193. Encuentre el valor del parámetro  $\theta$  de tal forma que la varianza de la siguiente distribución sea uno.

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

$$\text{Respuesta: } \theta = \sqrt{2}.$$

194. Sea  $X$  una variable aleatoria con función de densidad como aparece a continuación, en donde  $a$  y  $b$  son dos constantes.

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

Determine el valor de las constantes  $a$  y  $b$  de tal forma que:

- a) la esperanza sea mínima.  
b) la varianza sea mínima.
195. Demuestre las siguientes propiedades de la varianza:
- a)  $\text{Var}(X) \leq E(X^2)$ .  
b)  $\text{Var}(a - X) = \text{Var}(X)$ ,  $a$  constante.  
c)  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ ,  $a, b$  constantes.  
d)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + E[(X - E(X))(Y - E(Y))]$ .
196. Sea  $X$  una variable aleatoria con media  $\mu$  y varianza finita  $\sigma^2$ . Defina la función real  $g(x) = E(X - x)^2$ . Demuestre que:
- a)  $g(x) = \sigma^2 + (\mu - x)^2$ .  
b)  $g(x)$  tiene un mínimo en  $x = \mu$  y que ese valor mínimo es  $\sigma^2$ .
197. Demuestre o proporcione un contraejemplo:
- a)  $\text{Var}(\text{Var}(X)) = 0$ .  
b)  $\text{Var}(E(X)) = E(X)$ .  
c)  $E(\text{Var}(X)) = \text{Var}(X)$ .  
d)  $\text{Var}(X - Y) = \text{Var}(X) - \text{Var}(Y)$ .  
e) Si  $E(X)$  existe entonces  $\text{Var}(X)$  existe.  
f) Si  $\text{Var}(X)$  existe entonces  $E(X)$  existe.  
g) Si  $\text{Var}(X) = 0$  entonces  $X = 0$ .  
h) Si  $\text{Var}(X) = \text{Var}(Y)$  entonces  $X = Y$ .
198. Sean  $X_1, X_2, \dots, X_n$  v.a.s independientes e idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$ . La **media muestral** se define como la v.a.

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k.$$

Demuestre que:

- a)  $E(\bar{X}) = \mu$ .

$$b) \operatorname{Var}(\bar{X}) = \sigma^2/n.$$

199. Sean  $X$  y  $Y$  dos variables aleatorias con varianza finita. Demuestre que

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) \Leftrightarrow E(XY) = E(X)E(Y).$$

## 2.8. Momentos

Para cada número natural  $n$  se define el  $n$ -ésimo **momento** de una variable aleatoria  $X$  como el número  $E(X^n)$ , suponiendo que tal esperanza existe. Así, tenemos la siguiente definición.

**Definición 2.8** Los momentos de una variable aleatoria  $X$  son la colección de números:

$$E(X), E(X^2), E(X^3), \dots$$

correspondientes al primer momento, segundo momento,  $\dots$ , cuando tales cantidades existen.

Para variables aleatorias discretas el  $n$ -ésimo momento se calcula como sigue

$$E(X^n) = \sum_x x^n f(x),$$

mientras que para variables aleatorias continuas la fórmula es

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx.$$

Suponiendo su existencia, cada uno de estos números representa una característica de la variable aleatoria o de su distribución. Por ejemplo, el primer momento es el valor promedio o la esperanza de la variable aleatoria. Recordando la fórmula  $\operatorname{Var}(X) = E(X^2) - E^2(X)$ , podemos decir que la varianza es el segundo momento menos el primer momento al cuadrado, de este modo

el segundo momento está relacionado con la dispersión de los valores que toma la variable aleatoria. El tercer momento está relacionado con la simetría de la correspondiente distribución de probabilidad. En general no se conoce una interpretación para cada uno de los momentos de una variable aleatoria en el mismo sentido que no se conoce una interpretación para cada una de las derivadas de una función infinitamente diferenciable.

Por otro lado también debemos mencionar que los momentos pueden no existir y que en caso de que existan, en general no es de esperarse que pueda encontrarse una expresión compacta para el  $n$ -ésimo momento de una variable aleatoria dada. En la sección 2.11 definiremos la **función generadora de momentos**, la cual nos permitirá calcular los momentos de una variable aleatoria de una forma alternativa al cálculo de la suma o integral de la definición.

En el así llamado **problema de los momentos** se plantea encontrar condiciones bajo las cuales a partir del conjunto de todos los momentos de una variable aleatoria se puede identificar de manera única su distribución de probabilidad.

**Ejemplo 2.24** Considere una variable aleatoria continua  $X$  con función de densidad como aparece abajo. No es difícil comprobar que el primer momento es  $E(X) = 0$  y el segundo momento es  $E(X^2) = 1/6$ .

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

•

Se pueden definir también los siguientes momentos para una variable aleatoria:

$E(X - \mu)^n$	$n$ -ésimo momento central
$E X ^n$	$n$ -ésimo momento absoluto
$E X - \mu ^n$	$n$ -ésimo momento absoluto central
$E(X - c)^n$	$n$ -ésimo momento generalizado ( $c$ constante)

## Ejercicios

200. Demuestre que el  $n$ -ésimo momento de una variable aleatoria continua  $X$  con función de densidad

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

es

$$E(X^n) = \frac{2^{n+1}}{n+2}.$$

201. Sea  $X$  una variable aleatoria continua con media  $\mu$  y con función de densidad

$$f(x) = \begin{cases} x & \text{si } 0 < x < 1, \\ 2 - x & \text{si } 1 \leq x < 2, \\ 0 & \text{otro caso.} \end{cases}$$

Demuestre que el  $n$ -ésimo momento central de  $X$  es

$$E(X - \mu)^n = \frac{1 + (-1)^n}{(n+1)(n+2)}.$$

202. Sea  $X$  una variable aleatoria con distribución uniforme en el intervalo  $[a, b]$  y sea  $n$  cualquier número natural. Demuestre que:

$$a) E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

$$b) E(X - \mu)^n = \begin{cases} 0 & \text{si } n \text{ es impar,} \\ \frac{(b-a)^n}{(n+1)2^n} & \text{si } n \text{ es par.} \end{cases}$$

203. **Función de probabilidad simétrica.** Una función de probabilidad  $f(x)$  es simétrica respecto de  $a$  si  $f(a+x) = f(a-x)$  para cualquier número real  $x$ . Sea  $X$  una variable aleatoria con esperanza finita y con función de probabilidad simétrica respecto de  $a$ . Demuestre que:

$$a) E(X) = a.$$

$$b) E[(X - a)^n] = 0 \text{ para } n = 1, 3, 5, \dots$$

204. Sea  $X$  una variable aleatoria con segundo momento finito. Demuestre que

$$0 \leq E^2(X) \leq E(X^2).$$

205. Sea  $X$  una variable aleatoria con tercer momento finito. Demuestre o proporcione un contraejemplo para la siguiente afirmación:

$$E(X) \leq E(X^2) \leq E(X^3).$$

206. Demuestre que toda variable aleatoria acotada tiene todos sus momentos finitos.

## 2.9. Cuantiles

Los cuantiles son otras características numéricas de las distribuciones de probabilidad y se definen de la siguiente forma: sabemos que toda función de distribución  $F(x)$  crece de manera continua o a través de saltos, si  $p \in (0, 1]$  es una cierta probabilidad, entonces al valor más pequeño  $x$  tal que  $F(x)$  alcanza el nivel  $p$  o un nivel superior se le llama el **cuantil- $p$**  de la distribución y se le denota por  $c_p$ . Así, tenemos la siguiente definición.

**Definición 2.9** Sea  $p \in (0, 1]$ . El cuantil- $p$  de una v.a. o de su función de distribución  $F(x)$  es el número más pequeño  $c_p$ , si existe, tal que

$$F(c_p) \geq p. \quad (2.17)$$

En otras palabras,  $c_p$  es tal que la función de distribución acumula por lo menos una probabilidad  $p$  hasta ese valor. Cuando la función de distribución es continua, la desigualdad (2.17) se reduce a la igualdad:

$$F(c_p) = p.$$

Véase la Figura 2.10. Se acostumbra utilizar las siguientes expresiones: “ $c_p$  es el cuantil de orden  $p$ ”, o “ $c_p$  es el cuantil al  $100p\%$ ” de la distribución.

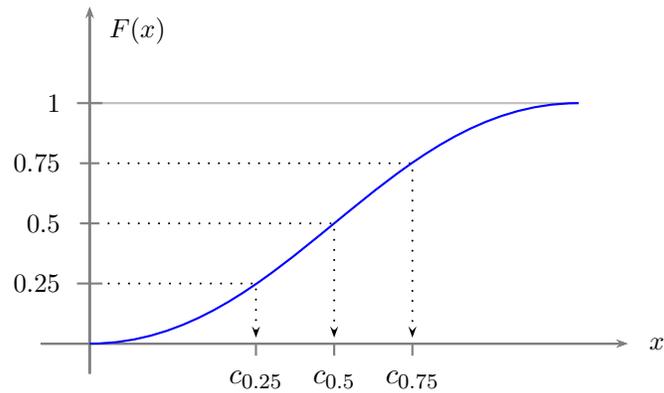


Figura 2.10

Así, por ejemplo,

$c_{0.1}$  es el cuantil al 10 %  
 $c_{0.2}$  es el cuantil al 20 %  
 ... ..

Los cuantiles son usados con frecuencia en algunos procedimientos de la estadística. En particular a las cantidades  $c_{0.25}$ ,  $c_{0.5}$ ,  $c_{0.75}$  y  $c_1$  se les llama **cuantiles** de la distribución. Más particularmente, al cuantil al 50% se le llama **mediana** de la distribución.

**Ejemplo 2.25** (Caso discreto) Sea  $X$  con función de distribución  $F(x)$  como se muestra en la Figura 2.11. Los cuatro cuantiles de esta distribución son:

$$\begin{aligned} c_{0.25} &= 2, & c_{0.75} &= 4, \\ c_{0.5} &= 3, & c_1 &= 5. \end{aligned}$$

•

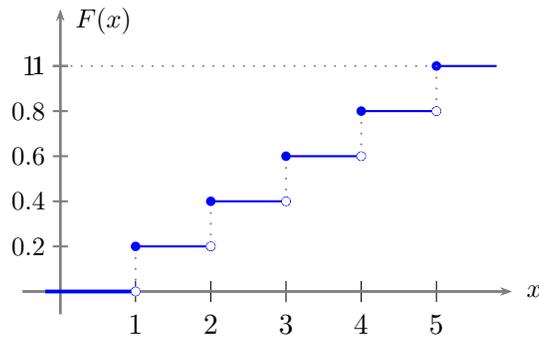


Figura 2.11

## Ejercicios

207. Calcule los cuantiles al 70 %, 80 % y 90 % para una variable aleatoria con función de probabilidad:

a) 

$x$	0	1	2
$f(x)$	1/2	1/4	1/4

b) 

$x$	-2	-1	0	1	2
$f(x)$	1/10	5/10	1/10	2/10	1/10

c) 

$x$	1	2	...	$n$	...
$f(x)$	1/2	1/4	...	1/2 <sup>n</sup>	...

208. Calcule los cuatro cuartiles de la función de distribución:

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

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

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

209. Encuentre la distribución de la variable aleatoria discreta  $X$  con únicamente dos posibles valores y tal que  $c_1 = 4$  y  $c_{0.3} = 2$ .

## 2.10. Función generadora de probabilidad

Sea  $X$  una variable aleatoria discreta con posibles valores dentro del conjunto  $\{0, 1, 2, \dots\}$ . Para este tipo de variables aleatorias vamos a asociar otra función equivalente a la función de probabilidad y a la función de distribución.

**Definición 2.10** Sea  $X$  una variable aleatoria discreta con posibles valores dentro del conjunto  $\{0, 1, 2, \dots\}$ . A la función  $G(t)$  definida como aparece abajo se le llama la **función generadora de probabilidad** de  $X$ ,

$$G(t) = E(t^X) = \sum_{x=0}^{\infty} t^x P(X = x). \quad (2.18)$$

Observe que dicha función está definida por lo menos para valores reales de  $t$  dentro del intervalo  $[-1, 1]$ , pues en tal caso la suma que aparece en (2.18) es convergente. En forma breve a esta función se le escribe como **f.g.p.** La letra  $G$  proviene del término “generadora” y para indicar que la variable aleatoria  $X$  es la asociada se le escribe también como  $G_X(t)$ . Así, la función  $G(t)$  se define como una serie de potencias en  $t$  con coeficientes dados por los valores de la función de probabilidad. Estos coeficientes pueden reconstruirse nuevamente a partir de la expresión de la función  $G(t)$  derivando y evaluando en cero, es decir, no es complicado demostrar que

$$P(X = x) = \frac{1}{x!} G^{(x)}(0), \quad x = 0, 1, \dots$$

en donde  $G^{(x)}(t)$  denota la derivada de orden  $x$  de  $G(t)$ . Esto justifica el nombre dado a esta función pues a partir de ella se pueden ir generando las probabilidades de que la variable aleatoria tome sus distintos valores. De esta

manera la f.g.p. proporciona una representación alterna de la información dada por la función de probabilidad. Véase el Ejercicio 210 para una breve lista de algunas otras propiedades de la f.g.p.

**Ejemplo 2.26** Considere la variable discreta  $X$  con función de probabilidad

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

En este caso la variable no toma valores enteros a partir de cero sino a partir de uno. Entonces

$$\begin{aligned} G(t) &= \sum_{x=1}^{\infty} t^x (1/2)^x \\ &= \sum_{x=1}^{\infty} (t/2)^x \\ &= \frac{t}{2-t} \quad \text{si } |t| < 2. \end{aligned}$$

Puede comprobarse que  $G^{(x)}(t) = 2x!(2-t)^{-x-1}$  y por lo tanto,

$$\frac{1}{x!} G^{(x)}(0) = 1/2^x = P(X = x).$$

•

El siguiente resultado nos provee de una fórmula para encontrar los momentos de una variable aleatoria a partir de su f.g.p. suponiendo la existencia de estos momentos. Para comprender mejor el enunciado debemos recordar que la expresión  $G(1-)$  se define como el límite de la función  $G(t)$  cuando  $t$  se aproxima al valor 1 por la izquierda.

**Proposición 2.9** Sea  $X$  discreta con función generadora de probabilidad  $G(t)$ . Si el  $k$ -ésimo momento de  $X$  existe entonces

$$G^{(k)}(1-) = E(X(X-1)\cdots(X-k+1)).$$

**Demostración.** Derivando  $k$  veces la serie de potencias (2.18) tenemos que

$$G^{(k)}(t) = \sum_{x=k}^{\infty} x(x-1)\cdots(x-k+1)t^{x-k}P(X=x).$$

Ahora se toma el límite cuando  $t \nearrow 1$ . El lema de Abel permite el intercambio de este límite con la suma infinita obteniéndose así el resultado anunciado, es decir,

$$\begin{aligned} G^{(k)}(1-) &= \sum_{x=k}^{\infty} x(x-1)\cdots(x-k+1)P(X=x) \\ &= E(X(X-1)\cdots(X-k+1)). \end{aligned}$$

•

El siguiente resultado es bastante útil en las aplicaciones de la f.g.p. y establece que la f.g.p. de la suma de dos variables independientes es el producto de las f.g.p.

**Proposición 2.10** Sean  $X$  y  $Y$  discretas e independientes con funciones generadoras de probabilidad  $G_X(t)$  y  $G_Y(t)$ . Entonces

$$G_{X+Y}(t) = G_X(t)G_Y(t).$$

**Demostración.** Usando la hipótesis de independencia tenemos que

$$\begin{aligned} G_{X+Y}(t) &= E(t^{X+Y}) \\ &= E(t^X t^Y) \\ &= E(t^X)E(t^Y) \\ &= G_X(t)G_Y(t). \end{aligned}$$

•

Es claro a partir de la definición que dos variables aleatorias con la misma distribución de probabilidad tienen asociada la misma f.g.p. Demostraremos

a continuación que la relación es uno a uno, es decir, si se tienen dos variables aleatorias con la misma f.g.p. entonces éstas tienen la misma distribución de probabilidad. Este resultado es muy importante pues establece que la f.g.p. caracteriza de manera única a la distribución de probabilidad de la variable aleatoria

**Proposición 2.11 (Caracterización).** Sean  $X$  y  $Y$  dos variables aleatorias discretas con el mismo conjunto de valores  $\{0, 1, 2, \dots\}$  y con funciones generadoras de probabilidad  $G_X(t)$  y  $G_Y(t)$ , tales que  $G_X(t) = G_Y(t)$  en un intervalo no trivial alrededor del cero. Entonces  $X$  y  $Y$  tienen la misma distribución de probabilidad.

**Demostración.** Supongamos que  $G_X(t) = G_Y(t)$  en un intervalo no trivial  $(-s, s)$  alrededor del cero. Esto significa que las dos series de potencias son idénticas en dicho intervalo. Substrayendo una serie de otra se obtiene una serie de potencias idénticamente cero, es decir,

$$\sum_{x=0}^{\infty} [P(X = x) - P(Y = x)] t^x = 0, \quad \text{para cada } t \in (-s, s).$$

Esto sólo es posible cuando los coeficientes de la serie son todos cero, es decir, para cada  $x = 0, 1, \dots$  se tiene que

$$P(X = x) = P(Y = x).$$

•

**Ejemplo 2.27** Se dice que la v.a.  $X$  tiene distribución Poisson de parámetro  $\lambda > 0$  si su función de probabilidad está dada por

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

Esto es, para cualquier valor real del parámetro  $\lambda > 0$  esta función es una función de probabilidad. En el siguiente capítulo estudiaremos con más

detalle esta distribución. Calcularemos a continuación su f.g.p.

$$\begin{aligned}
 G(t) &= E(t^X) \\
 &= \sum_{x=0}^{\infty} t^x e^{-\lambda} \frac{\lambda^x}{x!} \\
 &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(t\lambda)^x}{x!} \\
 &= e^{-\lambda} e^{t\lambda} \\
 &= e^{-\lambda(1-t)}.
 \end{aligned}$$

Así, debido a lo demostrado antes relativo a la correspondencia uno a uno entre las distribuciones de probabilidad y las f.g.p. sabemos que esta función es la f.g.p. de la distribución Poisson y que cualquier v.a. discreta con f.g.p. de esta forma tiene distribución Poisson. Usaremos este hecho para demostrar con facilidad que la suma de dos variables aleatorias independientes con distribución Poisson tiene nuevamente distribución Poisson. Sean  $X$  y  $Y$  dos variables aleatorias independientes con distribución Poisson de parámetros  $\lambda_1$  y  $\lambda_2$ , respectivamente. Entonces, por la independencia,

$$\begin{aligned}
 G_{X+Y}(t) &= G_X(t) G_Y(t) \\
 &= e^{-\lambda_1(1-t)} e^{-\lambda_2(1-t)} \\
 &= e^{-(\lambda_1+\lambda_2)(1-t)}.
 \end{aligned}$$

Observe que esta última expresión tiene la forma de la f.g.p. de la distribución Poisson sólo que en lugar del parámetro  $\lambda$  aparece la expresión  $\lambda_1 + \lambda_2$ . Esto indica que  $X + Y$  tiene distribución Poisson de parámetro  $\lambda_1 + \lambda_2$ . •

En los siguientes capítulos haremos uso de la f.g.p. y sus propiedades para caracterizar a algunas distribuciones de probabilidad específicas.

## Ejercicios

210. Sea  $X$  una v.a. discreta con valores en el conjunto  $\{0, 1, \dots\}$  y con f.g.p.  $G(t)$ . Demuestre las siguientes propiedades de la función generadora

de probabilidad. Recuerde que  $G(1-)$  se define como el límite de la función  $G(t)$  cuando  $t$  se aproxima al valor 1 por la izquierda. Véase además el lema de Abel que aparece en el apéndice.

a)  $G(1-) = 1$ .

b) Si  $X$  tiene esperanza finita, entonces

$$G^{(1)}(1-) = E(X).$$

c) Si  $X$  tiene varianza finita, entonces

$$G^{(2)}(1-) + G^{(1)}(1-) - [G^{(1)}(1-)]^2 = \text{Var}(X).$$

211. Sea  $X$  una variable aleatoria con función de probabilidad como aparece abajo. Demuestre que la f.g.p. de  $X$  es  $G(t) = (1 - t^n)/(n(1 - t))$ .

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

## 2.11. Función generadora de momentos

Otra función bastante útil que puede calcularse para algunas variables aleatorias, ahora incluyendo por igual el caso discreto y continuo, y que está relacionada con los momentos de la variable aleatoria es la que se define a continuación.

**Definición 2.11** La **función generadora de momentos** de una variable aleatoria discreta o continua  $X$  es la función  $M(t)$  definida como sigue:

$$M(t) = E(e^{tX}),$$

para valores reales de  $t$  en donde esta esperanza existe.

En forma breve se le escribe como **f.g.m.** La letra  $M$  corresponde al término “momentos” y en breve justificaremos su relación con los momentos de la

variable aleatoria. Cuando sea necesario especificar la variable aleatoria en cuestión se le escribe también como  $M_X(t)$ . Así, en el caso discreto esta función se calcula como

$$M(t) = \sum_x e^{tx} P(X = x),$$

y en el caso continuo

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

Hacemos énfasis en que la suma o integral anteriores pueden no ser convergentes para ningún valor de  $t$  distinto de cero y en tales casos decimos que la variable aleatoria no tiene función generadora de momentos finita. Se puede demostrar que cuando existe un intervalo no trivial  $(-s, s)$  alrededor del cero en donde la f.g.m. existe, entonces todos los momentos de  $X$  existen y la f.g.m. adquiere la forma de la siguiente serie de potencias:

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n). \quad (2.19)$$

En consecuencia,  $M(t)$  tiene derivadas continuas de cualquier orden en el intervalo  $(-s, s)$  y en consecuencia tenemos el siguiente resultado que justifica el nombre para esta función.

**Proposición 2.12** Sea  $X$  una v.a. con función generadora de momentos  $M(t)$  finita en un intervalo no trivial alrededor del cero. Entonces

$$M^{(n)}(0) = E(X^n).$$

**Demostración.** A partir de la expansión (2.19) es inmediato comprobar que para cualquier entero  $n \geq 0$ ,

$$M^{(n)}(0) = E(X^n).$$

Es decir, los momentos de  $X$  se encuentran derivando la f.g.m. y evaluando en  $t = 0$ . •

Demostraremos ahora que la f.g.m. de la suma de dos variables independientes es el producto de las f.g.m.

**Proposición 2.13** Sean  $X$  y  $Y$  independientes con funciones generadoras de momentos  $M_X(t)$  y  $M_Y(t)$ . Entonces

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

**Demostración.** Usando la hipótesis de independencia tenemos que

$$\begin{aligned} M_{X+Y}(t) &= E(e^{t(X+Y)}) \\ &= E(e^{tX} e^{tY}) \\ &= E(e^{tX}) E(e^{tY}) \\ &= M_X(t) M_Y(t). \end{aligned}$$

•

La f.g.m. también tiene la propiedad de caracterizar a la distribución de probabilidad de manera única. Este es el contenido del siguiente resultado cuya demostración no es sencilla y la omitiremos.

**Proposición 2.14 (Caracterización).** Sean  $X$  y  $Y$  dos variables aleatorias con f.g.m.  $M_X(t)$  y  $M_Y(t)$  las cuales coinciden en un intervalo no trivial alrededor del cero. Entonces  $X$  y  $Y$  tienen la misma distribución de probabilidad.

Más aún, las funciones generadoras de momentos cumplen con la siguiente propiedad importante que usaremos en el último capítulo para demostrar algunos teoremas límite.

**Teorema 2.1 (Continuidad de la f.g.m.).** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias tal que la  $X_n$  tiene f.g.m.  $M_{X_n}(t)$  y cuando  $n \rightarrow \infty$ ,

$$M_{X_n}(t) \rightarrow M_X(t), \quad t \in (-s, s),$$

para alguna v.a.  $X$  con f.g.m.  $M_X(t)$ . Entonces la sucesión de funciones de distribución  $F_{X_n}(x)$  converge a la función de distribución de la variable  $X$  en cada punto de continuidad  $x$  de  $F_X(x)$ , es decir, en tales puntos y cuando  $n \rightarrow \infty$ ,

$$F_{X_n}(x) \rightarrow F_X(x). \quad (2.20)$$

A la propiedad establecida en (2.20) la llamaremos convergencia en distribución de la sucesión de variables  $X_n$  a la variable  $X$ , y esto lo escribiremos como  $X_n \xrightarrow{d} X$ . Las demostraciones de estos dos últimos resultados pueden encontrarse en el texto de Gut [7].

## Ejercicios

212. Sea  $X$  una variable aleatoria discreta con función de probabilidad como aparece abajo. Encuentre la función generadora de momentos de  $X$  y a partir de ella calcule la media y la varianza de  $X$ .

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

Respuesta:  $M(t) = \frac{e^t}{2 - e^t}, \quad t < \ln 2.$

213. Sea  $X$  una variable aleatoria continua con función de densidad como aparece abajo. Encuentre la función generadora de momentos de  $X$  y a partir de ella calcule la media y la varianza de  $X$ .

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

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

214. Sea  $X$  una variable aleatoria con f.g.m.  $M_X(t)$  y sean  $a$  y  $b$  dos constantes. Demuestre que

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

215. **No existencia de la f.g.m.** Se dice que la variable aleatoria  $X$  tiene una distribución  $t$  con  $n = 1$  grados de libertad si su función de densidad es como aparece abajo. Demuestre que para esta distribución no existe su f.g.m.

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

## Capítulo 3

# Distribuciones de probabilidad

Estudiaremos ahora algunas distribuciones de probabilidad particulares. Las distribuciones que mencionaremos tienen un nombre particular adquirido ya sea debido a la situación en la que surge, o bien debido al nombre de su descubridor o a la persona que inicialmente la utilizó en alguna aplicación importante. Empezaremos con las distribuciones de tipo discreto y continuaremos después con las de tipo continuo. Principalmente en este último caso omitiremos especificar el experimento aleatorio y el espacio de probabilidad en donde pueden definirse estas variables aleatorias y sus distribuciones.

**Nota importante.** El lector debe siempre recordar que no existe homogeneidad en la literatura acerca de la forma de escribir los parámetros de una distribución de probabilidad dada. Por lo tanto se debe tener siempre cuidado al comparar fórmulas y resultados de una fuente bibliográfica a otra y también verificar la forma en la que los parámetros de una distribución particular son usados en sus implementaciones en los distintos lenguajes de programación y paquetes computacionales.

### 3.1. Distribución uniforme discreta

Decimos que una variable aleatoria  $X$  tiene una distribución uniforme discreta sobre el conjunto de  $n$  números  $\{x_1, \dots, x_n\}$  si la probabilidad de que  $X$  tome cualquiera de estos valores es constante  $1/n$ . Esta distribución surge

en espacios de probabilidad equiprobables, esto es, en situaciones en donde tenemos  $n$  resultados diferentes y todos ellos tienen la misma probabilidad de ocurrir. Los juegos de lotería son un ejemplo donde puede aplicarse esta distribución de probabilidad. Se escribe  $X \sim \text{unif}\{x_1, x_2, \dots, x_n\}$ , en donde el símbolo “ $\sim$ ” se lee “se distribuye como”. La función de probabilidad de esta variable aleatoria es

$$f(x) = \begin{cases} \frac{1}{n} & \text{si } x = x_1, x_2, \dots, x_n, \\ 0 & \text{otro caso.} \end{cases}$$

Es inmediato comprobar que la esperanza y la varianza para esta distribución se calculan del siguiente modo:

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

y

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

Algunas otras propiedades de esta distribución se encuentran en la sección de ejercicios.

**Ejemplo 3.1** La gráfica de la función de probabilidad de la distribución uniforme en el conjunto  $\{1, 2, 3, 4, 5\}$  aparece en la Figura 3.1, junto con la correspondiente función de distribución. Cada salto en la función de distribución es de tamaño  $1/5$ . La expresión completa de  $F(x)$  es la siguiente:

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1/5 & \text{si } 1 \leq x < 2, \\ 2/5 & \text{si } 2 \leq x < 3, \\ 3/5 & \text{si } 3 \leq x < 4, \\ 4/5 & \text{si } 4 \leq x < 5, \\ 1 & \text{si } x \geq 5. \end{cases}$$

•

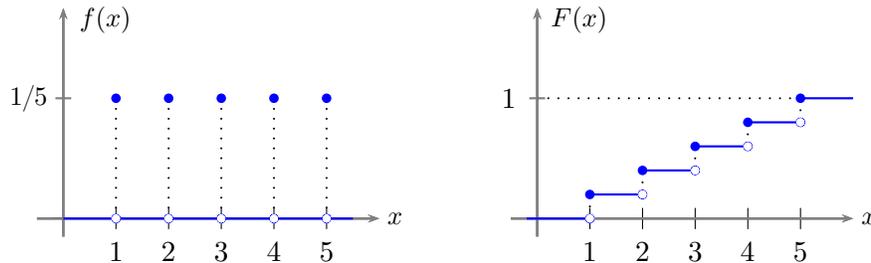


Figura 3.1

**Ejemplo 3.2** Al generar un número aleatorio en una computadora dentro del intervalo unitario  $[0, 1]$  y debido a que la precisión de la computadora es necesariamente finita, se obtienen siempre valores dentro de un conjunto finito de elementos. Por ejemplo, si la precisión de la computadora es de dos decimales, entonces sólo se pueden generar los números : 0.00, 0.01, 0.02, ..., 0.99, 1.00. La precisión de una computadora actual es claramente mucho mayor a la considerada pero siempre es finita y algún grado de imprecisión prevalece, es decir, en términos prácticos se tiene una distribución uniforme discreta al generar un valor al azar dentro del intervalo  $[0, 1]$ . •

**Ejemplo 3.3** En R se pueden generar valores al azar de la distribución uniforme discreta usando el siguiente comando:

```
# 15 valores al azar de la distribución unif{1,...,10}
> sample(1:10,15,replace=TRUE)
[1] 7 3 4 1 1 3 2 1 6 5 3 8 2 9 1
```

**Ejemplo 3.4** En Python se puede crear una lista de elementos y mediante la función predefinida `choice()` se escoge un elemento al azar en la lista con distribución uniforme. El código es el siguiente:

```

>>> import random
>>> conjunto=[1,2,3,4,5]
>>> random.choice(conjunto)
3

```

## Ejercicios

216. Sea  $X$  con distribución uniforme en el conjunto  $\{1, \dots, n\}$ . Demuestre que:

- a)  $E(X) = (n + 1)/2$ .
- b)  $E(X^2) = (n + 1)(2n + 1)/6$ .
- c)  $\text{Var}(X) = (n^2 - 1)/12$ .

217. Sea  $X$  con distribución  $\text{unif}\{0, 1\}$ . Demuestre que el  $n$ -ésimo momento de  $X$  es  $E(X^n) = 1/2$ .

218. Encuentre los cuatro cuartiles de la distribución  $\text{unif}\{1, \dots, 4n\}$ .

219. Sea  $X$  con distribución  $\text{unif}\{1, \dots, n\}$ . Demuestre que la f.g.p. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las fórmulas para  $E(X)$  y  $\text{Var}(X)$  que aparecen en el Ejercicio 216.

$$G(t) = \frac{t(1 - t^n)}{n(1 - t)}.$$

220. Sea  $X$  con distribución  $\text{unif}\{1, \dots, n\}$ . Demuestre que la f.g.m. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las fórmulas para  $E(X)$  y  $\text{Var}(X)$  que aparecen en el Ejercicio 216.

$$M(t) = \frac{e^t(1 - e^{nt})}{n(1 - e^t)}.$$

221. **Simulación.** Este es un mecanismo para generar valores al azar de una v.a con distribución  $\text{unif}\{x_1, \dots, x_n\}$  a partir de valores de una v.a. con distribución  $\text{unif}(0, 1)$ . Sea  $u$  un valor al azar con distribución

$\text{unif}(0, 1)$ , la cual estudiaremos mas adelante. Demuestre que la variable aleatoria  $X$  definida a continuación tiene distribución  $\text{unif}\{x_1, \dots, x_n\}$ .

$$X = \begin{cases} x_1 & \text{si } 0 < u \leq 1/n, \\ x_2 & \text{si } 1/n < u \leq 2/n, \\ \dots & \dots\dots \\ x_{n-1} & \text{si } (n-2)/n < u \leq (n-1)/n, \\ x_n & \text{si } (n-1)/n < u < 1. \end{cases}$$

222. Se escogen al azar y de manera independiente dos números  $a$  y  $b$  dentro del conjunto  $\{1, \dots, 10\}$ . Calcule la probabilidad de que:

- a)  $a$  y  $b$  coincidan.
- b)  $a$  sea menor a  $b$ .
- c)  $a$  sea mayor a  $b + 1$ .
- d)  $a$  y  $b$  difieran en por lo menos 2 unidades.

223. Sea  $X$  con distribución uniforme en el conjunto  $\{1, 2, 3, 4, 5\}$ . ¿Cuál es la probabilidad de que el área del rectángulo de lados  $X$  y  $6 - X$  sea mayor o igual a 8?

Respuesta:  $3/5$ .

224. Sean  $X$  y  $Y$  dos variables aleatorias independientes, ambas con distribución  $\text{unif}\{1, \dots, n\}$ . Encuentre una expresión para la función de probabilidad de la variable aleatoria:

- a)  $Z = \text{mín}\{X, Y\}$ .
- b)  $Z = \text{máx}\{X, Y\}$ .

## 3.2. Distribución Bernoulli

Un **ensayo Bernoulli** se define como aquel experimento aleatorio con únicamente dos posibles resultados llamados genéricamente: éxito y fracaso. Supondremos que las probabilidades de estos resultados son  $p$  y  $1 - p$  respectivamente. Si se define la variable aleatoria  $X$  como aquella función que

lleva el resultado éxito al número 1 y el resultado fracaso al número 0, entonces decimos que  $X$  tiene una distribución Bernoulli con parámetro  $p \in (0, 1)$  y escribimos  $X \sim \text{Ber}(p)$ . La función de probabilidad se puede escribir de la siguiente forma:

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

O bien de manera compacta,

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

La gráfica de esta función de probabilidad para  $p = 0.7$  aparece en la Figura 3.2, junto con la correspondiente función de distribución, la cual tiene la siguiente forma:

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

La función de probabilidad  $f(x)$  puede obtenerse en R usando el comando que se muestra a continuación. El nombre asignado al comando y sus argumentos serán justificados una vez que estudiemos la distribución binomial, pues resulta que la distribución Bernoulli es un caso particular de la distribución binomial.

```
# dbinom(x,1,p) evalúa f(x) de la dist. Ber(p)
> dbinom(0,1,0.7)
[1] 0.3
```

Para la función de distribución  $F(x)$  se usa el siguiente comando,

```
# pbinom(x,1,p) evalúa F(x) de la dist. Ber(p)
> pbinom(0.2,1,0.7)
[1] 0.3
```

Es inmediato verificar que

$$E(X) = p,$$

$$\text{Var}(X) = p(1 - p).$$

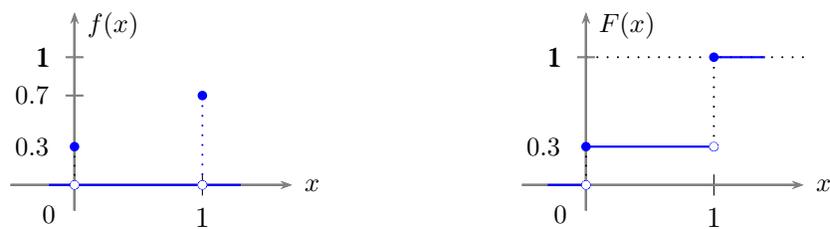


Figura 3.2

En la realización de todo experimento aleatorio siempre es posible preguntarnos por la ocurrencia o no ocurrencia de un evento cualquiera. Este es el esquema general donde surge esta distribución de probabilidad. La distribución Bernoulli es sencilla pero de muy amplia aplicación como veremos más adelante.

**Ejemplo 3.5** Sea  $\Omega$  el espacio muestral de un experimento aleatorio y sea  $A$  un evento con probabilidad  $p > 0$ . Sea  $X$  la variable aleatoria dada por

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Entonces  $X$  tiene distribución  $\text{Ber}(p)$ . A esta variable aleatoria  $X$  se le llama la **función indicadora** del evento  $A$  y se le denota también por  $1_A(\omega)$ . Así, al efectuar un ensayo del experimento aleatorio, la función indicadora señala cuando ocurre el evento  $A$  tomando el valor 1, e indica que no ha ocurrido el evento  $A$  tomando el valor 0. •

**Ejemplo 3.6** Considere el experimento aleatorio de lanzar una moneda al aire. Suponga que  $\omega_0$  y  $\omega_1$  son los dos resultados posibles, con probabilidades  $1 - p$  y  $p$ , respectivamente. Sea  $X$  la variable aleatoria dada por  $X(\omega_0) = 0$ , y  $X(\omega_1) = 1$ . Entonces  $X$  tiene distribución  $\text{Ber}(p)$ . ¿Puede usted encontrar la distribución de la variable  $Y = 1 - X$ ? •

**Simulación 3.1** En R se pueden generar  $k$  valores al azar de la distribución  $\text{Ber}(p)$  usando el comando que aparece en el siguiente recuadro. Modifique los valores de los parámetros y obtenga tantos valores al azar de esta distribución como desee.

```
# rbinom(k,1,p) genera k valores al azar de la dist. Ber(p)
> rbinom(25,1,0.7)
[1] 1 0 1 1 1 1 1 1 0 1 1 0 0 1 1 1 0 0 0 1 1 1 0 0 1
```

## Ejercicios

225. Sea  $X$  con distribución  $\text{Ber}(p)$  y sea  $n$  un número natural. Encuentre la distribución de las siguientes variables aleatorias:

- a)  $X^n$ .
- b)  $(1 - X)^n$ .
- c)  $|X - 1|^n$ .

Respuestas: a)  $\text{Ber}(p)$ . b)  $\text{Ber}(1 - p)$ . c)  $\text{Ber}(1 - p)$ .

226. Sea  $X$  con distribución  $\text{Ber}(p)$  y sean  $a$  y  $b$  dos constantes con  $a \neq 0$ . Defina la variable aleatoria  $Y = aX + b$ . Demuestre que:

- a) la función de probabilidad de  $Y$  es

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

- b)  $E(Y) = ap + b$ .
- c)  $\text{Var}(X) = a^2p(1 - p)$ .
- d)  $E(Y^n) = b^n(1 - p) + (a + b)^np, \quad n \geq 1$ .

227. Sea  $X$  con distribución  $\text{Ber}(p)$ . Demuestre que el  $n$ -ésimo momento de  $X$  es  $E(X^n) = p$ .
228. Sea  $X$  con distribución  $\text{Ber}(p)$ . Encuentre el valor de  $p$  tal que la varianza de  $X$  es máxima.
229. Encuentre los cuatro cuartiles de la distribución  $\text{Ber}(p)$  con  $p = 1/2$ .
230. Sea  $X$  con distribución  $\text{Ber}(p)$ . Demuestre que la f.g.p. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = p$  y  $\text{Var}(X) = p(1 - p)$ .

$$G(t) = 1 - p + pt.$$

231. Sea  $X$  con distribución  $\text{Ber}(p)$ . Demuestre que la f.g.m. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = p$  y  $\text{Var}(X) = p(1 - p)$ .

$$M(t) = 1 - p + pe^t.$$

232. **Simulación.** Este es un mecanismo para generar valores al azar de una v.a. con distribución  $\text{Ber}(p)$  a partir de valores de una v.a. con distribución  $\text{unif}(0, 1)$ . Sea  $u$  un valor al azar con distribución  $\text{unif}(0, 1)$ . Demuestre que la variable aleatoria  $X$  definida a continuación tiene distribución  $\text{Ber}(p)$ .

$$X = \begin{cases} 0 & \text{si } 0 < u \leq 1 - p, \\ 1 & \text{si } 1 - p < u < 1. \end{cases}$$

233. Sean  $X$  y  $Y$  dos v.a.s independientes con idéntica distribución  $\text{Ber}(p)$ . Encuentre la distribución de:

- |              |                  |
|--------------|------------------|
| a) $X + Y$ . | d) $X(1 - Y)$ .  |
| b) $X - Y$ . | e) $X(1 - X)$ .  |
| c) $XY$ .    | f) $X + Y - 1$ . |

234. Sean  $X_1, \dots, X_n$  v.a.i.i.d. con distribución  $\text{Ber}(p)$ . Demuestre que

$$X_1 \cdots X_n \sim \text{Ber}(p^n).$$

### 3.3. Distribución binomial

Supongamos que efectuamos una serie de  $n$  ensayos independientes Bernoulli en donde la probabilidad de éxito en cada ensayo es  $p$ . Si denotamos por  $E$  el resultado éxito y por  $F$  el resultado fracaso, entonces el espacio muestral de este experimento consiste de todas las posibles sucesiones de longitud  $n$  de caracteres  $E$  y  $F$ . Así, el espacio muestral consiste de  $2^n$  elementos. Si ahora definimos la variable aleatoria  $X$  como aquella función que indica el número de éxitos en cada una de estas sucesiones, esto es,

$$\begin{aligned} X(EE \cdots EE) &= n, \\ X(FE \cdots EE) &= n - 1, \\ &\vdots \\ X(FF \cdots FF) &= 0, \end{aligned}$$

entonces tenemos que  $X$  puede tomar los valores  $0, 1, 2, \dots, n$  con las probabilidades dadas por la función de probabilidad:

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

Decimos entonces que  $X$  tiene una distribución binomial con parámetros  $n$  y  $p$ , y escribimos  $X \sim \text{bin}(n, p)$ . Esta expresión para la función de probabilidad puede obtenerse de la siguiente forma: la probabilidad de obtener  $x$  éxitos y  $n - x$  fracasos en  $n$  ensayos Bernoulli es, preliminarmente,

$$\underbrace{p \cdots p}_x \underbrace{(1-p) \cdots (1-p)}_{n-x} = p^x (1-p)^{n-x},$$

pero hemos colocado los  $x$  éxitos en los primeros ensayos cuando ello no ocurrirá necesariamente así. Las diferentes formas en que los  $x$  éxitos pueden distribuirse en los  $n$  ensayos está dada por el coeficiente binomial  $\binom{n}{x}$ . Al hacer la multiplicación de estas cantidades se obtiene la expresión anunciada. El cálculo de la función  $f(x)$  puede representar un reto desde el punto de vista numérico pues se ven involucradas varias multiplicaciones particularmente cuando  $n$  es grande. En R esta función de probabilidad se obtiene usando el siguiente comando:

```
# dbinom(x,n,p) evalúa f(x) de la dist. bin(n,p)
> dbinom(8,10,0.3)
[1] 0.001446701
```

**Ejemplo 3.7** Cuando el número de ensayos es  $n = 10$  y la probabilidad de éxito es  $p = 0.3$ , se puede calcular, por ejemplo:

$$P(X = 2) = \binom{10}{2} (0.3)^2 (0.7)^{10-2} = 0.2334,$$

y de manera análoga el resto de las probabilidades. La gráfica de esta función de probabilidad con los parámetros  $n$  y  $p$  indicados se muestra en la Figura 3.3.

•

La función de distribución  $F(x)$  se escribe simplemente como la suma de los valores  $f(u)$  para valores de  $u$  menores a iguales a  $x$ , pero esta fórmula no tiene una expresión compacta y por ello no la escribiremos. Los valores de esta función se pueden obtener en R de la siguiente forma:

```
# pbinom(x,n,p) evalúa F(x) de la dist. bin(n,p)
> pbinom(4,10,0.3)
[1] 0.8497317
```

Por otro lado, después de algunos cálculos puede demostrarse que para una variable  $X$  con distribución  $\text{bin}(n, p)$ ,

$$E(X) = np, \tag{3.1}$$

$$\text{Var}(X) = np(1 - p). \tag{3.2}$$

$x$	$f(x)$
0	0.02824752
1	0.1210608
2	0.2334744
3	0.2668279
4	0.2001209
5	0.1029193
6	0.03675691
7	0.009001692
8	0.001446701
9	0.000137781
10	0.0000059049

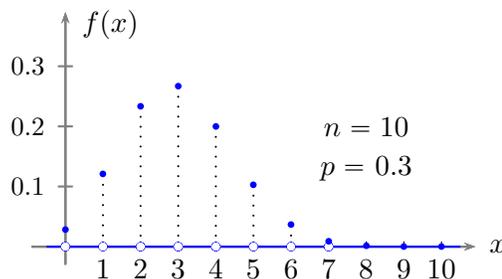


Figura 3.3

Cuando el parámetro  $n$  en la distribución  $\text{bin}(n, p)$  toma el valor 1 se obtiene la distribución  $\text{Ber}(p)$ . El siguiente resultado es muy útil y se puede demostrar por separado o bien considerarse como una consecuencia de la forma en la que se ha definido la distribución binomial.

**Proposición 3.1** Sean  $X_1, \dots, X_n$  variables aleatorias independientes, cada una con distribución  $\text{Ber}(p)$ . Entonces

$$X_1 + X_2 + \dots + X_n \sim \text{bin}(n, p). \quad (3.3)$$

Recíprocamente, toda v.a. con distribución  $\text{bin}(n, p)$  puede ser expresada como una suma de la forma anterior.

Así, cada sumando toma el valor 1 o 0 dependiendo de si el ensayo correspondiente fue éxito o fracaso, y la suma indica el número total de éxitos en los  $n$  ensayos. Aplicando las propiedades de la esperanza y la varianza en esta suma de variables aleatorias se pueden encontrar de forma más directa las expresiones (3.1) y (3.2).

**Simulación 3.2** La expresión (3.3) sugiere un mecanismo para generar valores al azar de la distribución  $\text{bin}(n, p)$ . Si se generan de manera independientes  $n$  valores al azar de la distribución  $\text{Ber}(p)$  y se suman estos valores, se obtiene un valor al azar de la distribución  $\text{bin}(n, p)$ . •

**Simulación 3.3** En R se pueden obtener valores al azar de la distribución  $\text{bin}(n, p)$  usando el comando que se muestra en el siguiente recuadro. Asigne valores a los parámetros correspondientes y genere tantos valores al azar como desee.

```
# rbinom(k,n,p) genera k valores al azar de la dist. bin(n,p)
> rbinom(25,10,0.3)
[1] 3 4 6 3 0 1 1 2 4 4 4 4 5 1 2 4 2 1 2 4 5 7 2 4 3
```

**Ejemplo 3.8** Un examen tiene diez preguntas y cada una tiene tres opciones como respuesta, siendo solamente una de ellas la correcta. Si un estudiante contesta cada pregunta al azar, ¿cuál es la probabilidad de que apruebe el examen?

Solución: si  $X$  denota el número de preguntas contestadas correctamente, entonces  $X$  tiene distribución  $\text{bin}(n, p)$  con  $n = 10$  y  $p = 1/3$ . Suponiendo que la calificación mínima aprobatoria es 6, entonces la respuesta es

$$P(X \geq 6) = \sum_{x=6}^{10} \binom{10}{x} (1/3)^x (2/3)^{10-x} = 0.07656353.$$

Esta probabilidad es sorprendentemente pequeña y por lo tanto la estrategia seguida por el estudiante para contestar el examen no parece ser la mejor. •

## Ejercicios

235. Sea  $X$  una variable aleatoria con distribución  $\text{bin}(n, p)$ . Use el teorema del binomio para demostrar que  $f(x)$  es efectivamente una función de probabilidad.

236. Sea  $X$  una variable aleatoria con distribución  $\text{bin}(n, p)$ . Encuentre los valores de los parámetros  $n$  y  $p$  cuando:

- a)  $E(X) = 6$  y  $\text{Var}(X) = 3$ .
- b)  $E(X) = 12$  y  $E(X^2) = 150$ .

Respuestas: a)  $n = 12$  y  $p = 1/2$ .  
b)  $n = 24$  y  $p = 1/2$ .

237. Sea  $f(x)$  la función de probabilidad de la distribución  $\text{bin}(n, p)$ . Demuestre que:

- a) **Fórmula iterativa:** para  $x$  entero tal que  $0 \leq x \leq n$ ,

$$f(x+1) = \frac{p}{(1-p)} \frac{(n-x)}{(x+1)} f(x).$$

- b)  $f(x)$  es creciente de  $x$  a  $x+1$  para valores enteros de  $x$  en el intervalo  $[0, np+p-1]$ .
- c)  $f(x)$  es decreciente de  $x$  a  $x+1$  para valores enteros de  $x$  en el intervalo  $[np+p-1, n]$ .
- d)  $f(x)$  tiene un máximo en  $x^*$  definido como el entero más pequeño mayor o igual a  $np+p-1$ , es decir,  $x^* = \lceil np+p-1 \rceil$ .
- e)  $f(x^*) = f(x^*+1)$  cuando  $x^*$  coincide con  $np+p-1$ . Por el inciso anterior, esto significa que  $f(x)$  tiene dos modas o máximos, uno en  $x^*$  y otro en  $x^*+1$ .

238. Usando directamente la definición de esperanza, demuestre que si  $X$  tiene distribución  $\text{bin}(n, p)$ , entonces:

- a)  $E(X) = np$ .
- b)  $E(X^2) = np(1-p+np)$ .
- c)  $\text{Var}(X) = np(1-p)$ .

239. Sea  $X$  con distribución  $\text{bin}(n, p)$ . Demuestre que la f.g.p. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = np$  y  $\text{Var}(X) = np(1-p)$ .

$$G(t) = (1 - p + pt)^n.$$

240. Sea  $X$  con distribución  $\text{bin}(n, p)$ . Demuestre que la f.g.m. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = np$  y  $\text{Var}(X) = np(1-p)$ .

$$M(t) = (1 - p + pe^t)^n.$$

241. Use la f.g.m. para demostrar la primera parte de la Proposición 3.1 de la página 176.
242. Sean  $X$  y  $Y$  independientes con distribución  $\text{bin}(n, p)$  y  $\text{bin}(m, p)$  respectivamente. Demuestre que la variable  $X + Y$  tiene distribución  $\text{bin}(n + m, p)$  siguiendo los siguientes tres métodos:

- Calculando directamente  $P(X + Y = k)$  para  $k = 0, 1, \dots, n + m$ .
- Usando la f.g.p.
- Usando la f.g.m.

243. Demuestre que si  $X$  tiene distribución  $\text{bin}(n, p)$ , entonces  $n - X$  tiene distribución  $\text{bin}(n, 1 - p)$ .
244. Considere un experimento aleatorio y sea  $A$  un evento con probabilidad estrictamente positiva. Suponga que se realizan  $n$  ensayos independientes de este experimento aleatorio y defina a  $X$  como el número de veces que se observa la ocurrencia del evento  $A$  en estos  $n$  ensayos. Demuestre que para cualquier entero fijo  $k \geq 1$ ,

$$\lim_{n \rightarrow \infty} P(X > k) = 1.$$

245. [Regularidades estadísticas](#). Escriba un programa en R que efectúe lo siguiente:
- Asigne un valor natural al parámetro  $n$  y una probabilidad al parámetro  $p$ .

- b) Genere 200 valores independientes al azar  $x_1, \dots, x_{200}$  de la distribución  $\text{bin}(n, p)$  y calcule los promedios parciales

$$s_m = \frac{1}{m} \sum_{k=1}^m x_k, \quad \text{para } m = 1, 2, \dots, 200.$$

- c) Grafique la función  $m \mapsto s_m$  y una los puntos con líneas rectas. Grafique también la línea horizontal  $y = np$ .

¿Qué puede decir del comportamiento de  $s_m$ ? Esta regularidad se presenta siempre para cualquier distribución con esperanza finita y se llama ley de los grandes números.

246. Un productor de semillas conoce por experiencia que el 10% de un gran lote de semillas no germinan. El productor vende sus semillas en paquetes de 200 semillas garantizando que por lo menos 185 de ellas germinarán. Calcule el porcentaje de paquetes que no cumplirán la garantía.
247. Se conoce que en una cierta población el 0.3% de las personas tienen un cierto tipo de accidente en un año dado cualquiera. Encuentre la probabilidad de que una compañía aseguradora tenga que indemnizar a más de 5 personas de los 1,500 asegurados que componen su cartera para este tipo de accidentes en un año.
248. Se realizan 10 lanzamientos sucesivos e independientes de un dado equilibrado y se nos informa que han aparecido por lo menos  $k$  unos,  $1 \leq k \leq 9$ . Calcule la probabilidad de que:
- exactamente  $k$  unos se hayan obtenido.
  - por lo menos  $k + 1$  unos se hayan obtenido.
  - a lo sumo  $k + 1$  unos se hayan obtenido.
  - Todos hayan sido unos.

### 3.4. Distribución geométrica

Supongamos ahora que tenemos una sucesión infinita de ensayos independientes Bernoulli, en cada uno de los cuales la probabilidad de éxito es  $p$ . Para cada una de estas sucesiones definimos la variable aleatoria  $X$  como el número de fracasos antes de obtener el primer éxito. Por ejemplo,

$$\begin{aligned} X(FEFEEFF \dots) &= 1, \\ X(EFFEEEE \dots) &= 0, \\ X(FFFEEFE \dots) &= 3. \end{aligned}$$

Observamos que  $X$  puede tomar los valores  $0, 1, 2, \dots$ . No es difícil darse cuenta que la probabilidad de que  $X$  tome el valor entero  $x \geq 0$  es  $p(1-p)^x$ . Decimos entonces que  $X$  tiene una distribución geométrica con parámetro  $p$  y escribimos  $X \sim \text{geo}(p)$  cuando su función de probabilidad es

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

La gráfica de esta función cuando  $p = 0.4$  se muestra en la Figura 3.4 y en R se pueden encontrar los valores de  $f(x)$  de la siguiente forma:

```
# dgeom(x,p) evalúa f(x) de la dist. geo(p)
> dgeom(5,0.4)
[1] 0.031104
```

El nombre de esta distribución proviene del hecho de que cuando escribimos la suma de todas las probabilidades obtenemos una suma geométrica. Efectuando las sumas parciales de la función de probabilidad se encuentra que la correspondiente función de distribución es:

$$F(x) = \sum_{u \leq x} f(u) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - (1-p)^1 & \text{si } 0 \leq x < 1, \\ 1 - (1-p)^2 & \text{si } 1 \leq x < 2, \\ \dots & \dots \\ 1 - (1-p)^{k+1} & \text{si } k \leq x < k+1, \\ \dots & \dots \end{cases}$$

$x$	$f(x)$
0	0.4
1	0.24
2	0.144
3	0.0864
4	0.05184
5	0.031104
6	0.0186624
7	0.01119744
8	0.006718464
9	0.004031078
...	...

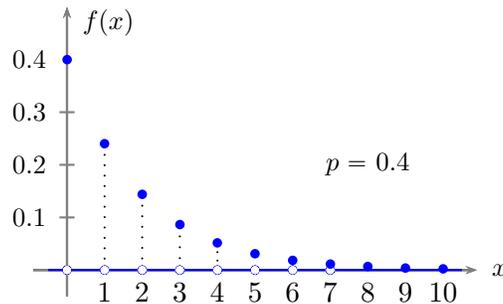


Figura 3.4

Estos valores pueden encontrarse en R usando el siguiente comando:

```
# pgeom(x,p) evalúa F(x) de la dist. geo(p)
> pgeom(5,0.4)
[1] 0.953344
```

Para esta distribución es posible además demostrar que

$$E(X) = \frac{1-p}{p},$$

$$\text{Var}(X) = \frac{1-p}{p^2}.$$

**Simulación 3.4** En R se pueden generar valores al azar de la distribución geométrica de la forma como se muestra en el siguiente recuadro. Asigne un valor al parámetro  $p$  y genere valores al azar de esta distribución.

```
# rgeom(k,p) genera k valores al azar de la dist. geo(p)
> rgeom(25,0.4)
[1] 0 1 1 0 7 5 0 4 1 0 0 0 0 0 0 0 2 1 1 0 1 0 0 1 0
```

•

**Ejemplo 3.9** La inspección sucesiva de artículos hasta encontrar uno defectuoso, posiblemente en un proceso de control de calidad, puede modelarse usando una distribución geométrica. •

**Ejemplo 3.10** Una persona participa cada semana con un boleto en un juego de lotería en donde la probabilidad de ganar el primer premio es  $p = 10^{-6} = 1/1,000,000$ . ¿Cuántos años en promedio debe esta persona participar en el juego hasta obtener el primer premio?

Solución: si  $X$  denota el número de participaciones en el juego antes de obtener el primer premio, entonces  $X$  tiene distribución geométrica de parámetro  $p = 10^{-6}$ . Y por lo tanto la esperanza de la variable  $X + 1$  es el número promedio de semanas que deben transcurrir hasta obtener el primer premio. Este número es

$$\frac{1-p}{p} + 1 = \frac{1}{p} = 10^6 = 1,000,000.$$

Lo que es equivalente a 19,230 años aproximadamente. •

## Ejercicios

249. Sea  $f(x)$  la función de probabilidad de la distribución  $\text{geo}(p)$ . Demuestre que:

- $f(x)$  es efectivamente una función de probabilidad.
- $f(x)$  es decreciente y por lo tanto tiene un máximo en  $x^* = 0$ .

250. Use la fórmula (2.15) del Ejercicio 176 para demostrar que si  $X$  tiene distribución  $\text{geo}(p)$  entonces

$$E(X) = \frac{1-p}{p}.$$

251. **Simulación.** Sea  $X_0, X_1, \dots$  una sucesión de v.a.s independientes con distribución  $\text{Ber}(p)$ . Defina

$$X = \min\{n \geq 0 : X_n = 1\}.$$

Demuestre que  $X$  tiene distribución  $\text{geo}(p)$ . Esto permite encontrar valores al azar de la distribución geométrica a partir de valores al azar de la distribución Bernoulli.

252. **f.g.p.** Sea  $X$  con distribución  $\text{geo}(p)$ . Demuestre que la f.g.p. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = (1-p)/p$  y  $\text{Var}(X) = (1-p)/p^2$ .

$$G(t) = \frac{p}{1 - (1-p)t}.$$

253. **f.g.m.** Sea  $X$  con distribución  $\text{geo}(p)$ . Demuestre que la f.g.m. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = (1-p)/p$  y  $\text{Var}(X) = (1-p)/p^2$ .

$$M(t) = \frac{p}{1 - (1-p)e^t}.$$

254. **Pérdida de memoria.** Sea  $X$  con distribución  $\text{geo}(p)$ . Demuestre que para cualesquiera enteros  $n, m \geq 0$ ,

$$P(X \geq n + m | X \geq m) = P(X \geq n).$$

255. Sean  $X$  y  $Y$  v.a.s independientes, ambas con distribución  $\text{geo}(p)$ . Demuestre que

$$P(X + Y = k) = \binom{k+1}{k} (1-p)^k p^2, \quad k = 0, 1, \dots$$

Compruebe que la expresión anterior corresponde efectivamente a una función de probabilidad. Esta es la distribución binomial negativa de parámetros  $(r, p)$  con  $r=2$  que veremos en la siguiente sección.

256. Dos personas lanzan alternativamente una moneda equilibrada. Se escoge una de las caras de la moneda y el primero que obtenga esa cara es el ganador. Encuentre la probabilidad de ganar de cada uno de los jugadores.

Respuesta:  $p_1 = 2/3$ ,  $p_2 = 1/3$ .

257. **La otra geométrica.** En ocasiones es necesario considerar el número de ensayos (no el de fracasos) antes del primer éxito en una sucesión de ensayos independientes Bernoulli. En este caso la variable es  $Y = 1 + X$  en donde  $X$  tiene distribución  $\text{geo}(p)$ , es decir, la distribución se desplaza hacia la derecha una unidad. Para la variable  $Y$  encuentre su función de probabilidad  $f_Y(y)$  y su función de distribución  $F_Y(y)$ . Demuestre además que:

a)  $E(Y) = 1/p$ .

b)  $\text{Var}(Y) = (1 - p)/p^2$ .

### 3.5. Distribución binomial negativa

Consideremos nuevamente la situación de observar los resultados de una sucesión infinita de ensayos independientes Bernoulli, en cada uno de los cuales la probabilidad de éxito es  $p$ . Sea  $r \geq 1$  un número entero. Definimos ahora a la variable aleatoria  $X$  como el número de fracasos antes de obtener el  $r$ -ésimo éxito. Decimos entonces que  $X$  tiene una distribución binomial negativa con parámetros  $r$  y  $p$ , y escribimos  $X \sim \text{bin neg}(r, p)$ . Es claro que la variable  $X$  puede tomar los valores  $0, 1, 2, \dots$  con las probabilidades dadas por la función de probabilidad:

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

En esta fórmula aparece el término  $p^r$  pues nos interesa observar  $r$  éxitos. Por otro lado, podemos tener un número variable  $x$  de fracasos, de ahí el término  $(1-p)^x$ . Finalmente el factor  $\binom{r+x-1}{x}$  indica los diferentes arreglos en los que los  $x$  fracasos y los  $r-1$  éxitos se encuentran distribuidos en

$r + x - 1$  ensayos. Observe que el  $r$ -ésimo éxito debe aparecer en el ensayo  $r + x$ . Como un ejercicio no trivial se deja al lector verificar que la función  $f(x)$  arriba indicada es efectivamente una función de probabilidad siguiendo la sugerencia que aparece en el Ejercicio 258. La gráfica de esta función aparece en la Figura 3.5 cuando los valores de los parámetros son  $r = 5$  y  $p = 0.5$ . Los valores de  $f(x)$  se pueden obtener en R de la siguiente forma:

```
# dnbinom(x,r,p) evalúa f(x) de la distribución bin. neg.(r,p)
> dnbinom(3,5,0.5)
[1] 0.1367188
```

$x$	$f(x)$
0	0.03125
1	0.078125
2	0.1171875
3	0.1367188
4	0.1367188
5	0.1230469
6	0.1025391
7	0.08056641
8	0.0604248
9	0.04364014
10	0.0305481
...	...

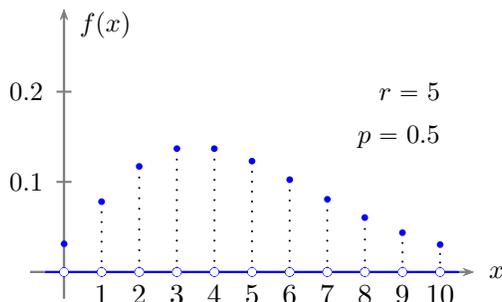


Figura 3.5

No existe una expresión compacta para la función de distribución  $F(x)$  de la distribución binomial negativa y por lo tanto no intentaremos escribirla. Sus valores se pueden encontrar en R usando el siguiente comando:

```
# pnbinom(x,r,p) evalúa F(x) de la distribución bin. neg.(r,p)
> pnbinom(7,5,0.5)
[1] 0.8061523
```

Es claro que la distribución binomial negativa es una generalización de la distribución geométrica. Esta última se obtiene tomando  $r = 1$ . Se puede

además demostrar que

$$E(X) = r \frac{1-p}{p},$$

$$\text{Var}(X) = r \frac{1-p}{p^2}.$$

Por otro lado, el coeficiente binomial puede extenderse para cualquier número real  $a$  y cualquier entero natural  $x$  de la siguiente manera:

$$\binom{a}{x} = \frac{a(a-1)\cdots(a-x+1)}{x!}. \quad (3.4)$$

Puede entonces demostrarse la siguiente identidad, de donde adquiere su nombre la distribución binomial negativa:

$$\binom{r+x-1}{x} = (-1)^x \binom{-r}{x}. \quad (3.5)$$

**Simulación 3.5** En R se pueden generar valores al azar de la distribución binomial negativa como se muestra en el siguiente recuadro. Asigne valores a los parámetros  $r$  y  $p$ , y genere valores al azar de esta distribución.

```
# rbinom(k,r,p) genera k valores al azar de la distribución
# bin. neg.(r,p)
> rbinom(25,5,0.5)
[1] 1 7 7 3 1 4 2 1 3 10 4 5 3 1 11 6 7 3 5 3 9 9 6 1 7
```

**Ejemplo 3.11** Se lanza repetidas veces una moneda honesta cuyos dos resultados son cara y cruz. ¿Cuál es la probabilidad de obtener la tercera cruz en el quinto lanzamiento?

Solución: sea  $X$  el número de caras (fracasos) antes de obtener la tercera cruz. Entonces  $X \sim \text{bin neg}(r, p)$  con  $r = 3$ ,  $p = 1/2$ . Nos preguntan  $P(X = 2)$ . Tenemos que

$$P(X = 2) = \binom{4}{2} (1/2)^5 = 6/32 = 0.1875.$$

## Ejercicios

258. Demuestre que la función de probabilidad de la distribución binomial negativa efectivamente es una función de probabilidad.

Sugerencia: Use la identidad (3.5) y la expansión  
 $(1+t)^a = \sum_{x=0}^{\infty} \binom{a}{x} t^x$ ,  $a \in \mathbb{R}$ ,  $|t| < 1$ .

259. Demuestre que el coeficiente binomial que aparece en la distribución binomial negativa se puede expresar de la siguiente forma:

$$\binom{r+x-1}{x} = (-1)^x \binom{-r}{x}.$$

260. Sea  $f(x)$  la función de probabilidad de la distribución bin neg( $r, p$ ). Demuestre que:

a) **Fórmula iterativa:** para  $x \geq 0$  entero,

$$f(x+1) = (1-p) \frac{x+r}{x+1} f(x)$$

b)  $f(x)$  es creciente de  $x$  a  $x+1$  para valores enteros de  $x$  dentro del intervalo  $[0, (r-1)(1-p)/p - 1]$ .

c)  $f(x)$  es decreciente de  $x$  a  $x+1$  para valores enteros de  $x$  dentro del intervalo  $[(r-1)(1-p)/p - 1, \infty)$ .

d)  $f(x)$  tiene un máximo en  $x^*$  definido como el entero más pequeño mayor o igual a  $(r-1)(1-p)/p - 1$ , es decir,  $x^* = \lceil (r-1)(1-p)/p - 1 \rceil$ .

e)  $f(x^*) = f(x^* + 1)$  cuando  $x^*$  coincide con  $(r-1)(1-p)/p - 1$ . Por el inciso anterior, esto significa que  $f(x)$  tiene dos modas o máximos, uno en  $x^*$  y otro en  $x^* + 1$ .

261. **Simulación.** Sea  $X_1, X_2, \dots$  una sucesión de v.a.s independientes con distribución  $\text{Ber}(p)$  y sea  $r \geq 1$  un entero. Defina

$$X = \text{mín}\{n \geq r : \sum_{k=1}^n X_k = r\} - r.$$

Demuestre que  $X$  tiene distribución bin. neg.  $(r, p)$ . Esto permite encontrar valores al azar de la distribución binomial negativa a partir de valores al azar de la distribución Bernoulli.

262. Sea  $X$  con distribución bin. neg.  $(r, p)$ . Demuestre que la f.g.p. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = r(1-p)/p$  y  $\text{Var}(X) = r(1-p)/p^2$ .

$$G(t) = \left( \frac{p}{1 - (1-p)t} \right)^r.$$

263. Sea  $X$  con distribución bin. neg.  $(r, p)$ . Demuestre que la f.g.m. de  $X$  está dada por la expresión que aparece abajo. A través de esta función encuentre nuevamente las expresiones  $E(X) = r(1-p)/p$  y  $\text{Var}(X) = r(1-p)/p^2$ .

$$M(t) = \left( \frac{p}{1 - (1-p)e^t} \right)^r.$$

264. Sean  $X$  y  $Y$  v.a.s independientes con distribución bin. neg.  $(r, p)$  y bin. neg.  $(s, p)$  respectivamente. Demuestre que la variable  $X + Y$  tiene distribución bin. neg.  $(r + s, p)$  siguiendo los siguientes tres métodos:

- a) Calculando directamente  $P(X + Y = k)$  para  $k = 0, 1, \dots$
- b) Usando la f.g.p.
- c) Usando la f.g.m.

265. **Muestreo.** Se desea encontrar a 20 personas que reúnan ciertas características para aplicarles un cuestionario. Si únicamente el 1% de la población cumple las características requeridas y suponiendo que se consulta al azar a las personas para determinar si son adecuadas para contestar el cuestionario, determine el número promedio de personas que se necesita consultar para encontrar a las 20 personas solicitadas.

### 3.6. Distribución hipergeométrica

Esta distribución de probabilidad surge en el contexto de la toma de una muestra de un conjunto de objetos de dos tipos. Supongamos que tenemos

$N$  objetos dentro de una caja, de los cuales  $K$  son de un primer tipo y  $N - K$  son de un segundo tipo. Véase la Figura 3.6. Los objetos del primer tipo pueden corresponder a artículos en buen estado y los del segundo tipo a artículos en mal estado, o bien a personas con una cierta característica y aquellas que no poseen dicha característica.

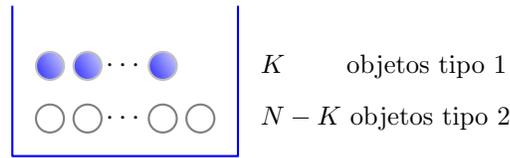


Figura 3.6

Supongamos que de esta caja tomamos al azar una muestra de tamaño  $n$  de tal forma que la selección es sin reemplazo y el orden de los objetos seleccionados no es relevante. Así, el espacio muestral de este experimento consiste de todos los posibles subconjuntos de tamaño  $n$  que se pueden obtener de esta colección de  $N$  objetos y su cardinalidad es entonces  $\binom{N}{n}$ . Si para cada subconjunto seleccionado se define la variable aleatoria  $X$  como el número de objetos seleccionados que son del primer tipo, entonces es claro que  $X$  puede tomar los valores  $0, 1, 2, \dots, n$ . Observe que  $X$  toma el valor  $n$  si y sólo si todos los objetos escogidos son del tipo 1, mientras que toma el valor  $0$  cuando todos los objetos escogidos son del tipo 2. Para que tales casos puedan ocurrir supondremos que el tamaño  $n$  de muestra es suficientemente pequeño de tal forma que:

$$n \leq \min \{K, N - K\}. \quad (3.6)$$

La probabilidad de que  $X$  tome un valor  $x$  está dada por la siguiente expresión:

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

Decimos entonces que  $X$  tiene una distribución hipergeométrica con parámetros  $N$ ,  $K$  y  $n$ , y escribimos  $X \sim \text{hipergeo}(N, K, n)$ . Para entender

la fórmula de la función de probabilidad de esta distribución observe que el término  $\binom{K}{x}$  establece las diferentes formas en que  $x$  objetos pueden escogerse de la colección de  $K$  objetos del tipo 1, mientras que el término  $\binom{N-K}{n-x}$  corresponde a las diferentes formas de escoger  $n - x$  objetos de los  $N - K$  objetos del tipo 2. Se usa entonces el principio multiplicativo para obtener el número total de muestras diferentes en donde  $x$  objetos son del primer tipo y  $n - x$  objetos son del segundo tipo. No es un ejercicio fácil verificar que esta función de probabilidad efectivamente lo es, pero puede realizarse usando la sugerencia que aparece en el Ejercicio 266. La gráfica de esta función de probabilidad para  $N = 20$ ,  $K = 7$  y  $n = 5$  aparece en la Figura 3.7.

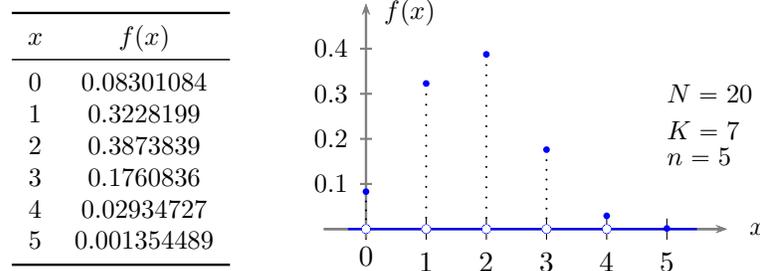


Figura 3.7

En R pueden obtenerse los valores de  $f(x)$  como se muestra en el recuadro siguiente. Observe con cuidado la diferencia en el orden y la forma de expresar los parámetros de esta distribución en R: después del argumento  $x$  se especifica el número de objetos  $K$  de tipo 1, después el número de objetos  $N - K$  de tipo 2 y finalmente se especifica el tamaño de la muestra  $n$ .

```
# dhyper(x,K,N-K,n) evalúa f(x) de la dist. hipergeo(N,K,n)
> dhyper(3,7,13,5)
[1] 0.1760836
```

Por otro lado, no presentaremos una fórmula para la función de distribución  $F(x)$  pues no tiene una expresión compacta sencilla, sin embargo sus valores pueden encontrarse usando R mediante el siguiente comando:

```
# phyper(x,K,N-K,n) evalúa  $F(x)$  de la dist. hipergeo( $N, K, n$ )
> phyper(3,7,13,5)
[1] 0.9692982
```

No es muy complicado comprobar que si  $X$  tiene distribución hipergeo( $N, K, n$ ), entonces

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

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

**Simulación 3.6** Mediante el siguiente comando en R pueden obtenerse valores al azar de la distribución hipergeométrica. Asigne usted valores a los parámetros  $N$ ,  $K$  y  $n$  como en el ejemplo y genere tantos valores de esta distribución como desee a través del valor de  $k$ .

```
# rhyper(k,K,N-K,n) genera  $k$  valores al azar de la distribución
# hipergeo( $N, K, n$ )
> rhyper(25,7,13,5)
[1] 3 2 3 2 3 1 3 2 1 1 1 3 1 3 3 0 4 3 4 3 1 1 3 2 1
```

## Ejercicios

266. Demuestre que la función de probabilidad de la distribución hipergeométrica efectivamente es una función de probabilidad.

Sugerencia:  $(a+b)^N = (a+b)^K (a+b)^{N-K}$ ,  
iguale coeficientes.

267. Sea  $X$  con distribución hipergeo( $N, K, n$ ). Demuestre que:

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

Sugerencia:  $x^2 = x(x - 1) + x$ .

268. Compruebe que la distribución hipergeo( $N, K, n$ ) se reduce a la distribución Ber( $p$ ) con  $p = K/N$  cuando  $n = 1$ .
269. Sea  $X$  con distribución hipergeo( $N, K, n$ ). Demuestre que la función de probabilidad de  $X$  converge a la función de probabilidad bin( $n, p$ ) cuando  $N \rightarrow \infty$  y  $K \rightarrow \infty$  de tal forma que  $K/N \rightarrow p$ .
270. Suponga que en conjunto se tienen  $N_1$  objetos de un primer tipo,  $N_2$  objetos de un segundo tipo y  $N_3$  objetos de un tercer tipo. Suponga que se extrae al azar un subconjunto de tamaño  $n$  de tal forma que  $1 \leq n \leq \min\{N_1, N_1 + N_2\}$ . Sea  $X$  el número de objetos del primer tipo contenidos en la muestra. Encuentra la distribución de  $X$ .

Solución: hipergeo( $N_1 + N_2 + N_3, N_1, n$ ).

271. Se pone a la venta un lote de 100 artículos de los cuales 10 son defectuosos. Un comprador extrae una muestra al azar de 5 artículos y decide que si encuentra 2 o más defectuosos, entonces no compra el lote. Calcule la probabilidad de que la compra se efectúe.

Solución: 0.9231433.

### 3.7. Distribución Poisson

Supongamos que deseamos observar el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado, por ejemplo, el número de clientes que llegan a un cajero automático durante la noche, o tal vez deseamos registrar el número de accidentes que ocurren en cierta avenida durante todo un día, o el número de reclamaciones que llegan a una compañía aseguradora. Para modelar este tipo de situaciones podemos definir la variable aleatoria  $X$  como el número de ocurrencia de este evento en el intervalo de tiempo dado. Es claro entonces que  $X$  puede tomar los valores  $0, 1, 2, \dots$ , y en principio no ponemos una cota superior para el número de observaciones del evento. Adicionalmente supongamos que conocemos la tasa media de ocurrencia del evento de interés, que denotamos por la letra  $\lambda$  (lambda). El parámetro  $\lambda$  es positivo y se interpreta como el número promedio de ocurrencias del evento por unidad de tiempo o espacio. La probabilidad de que

la variable aleatoria  $X$  tome un valor entero  $x \geq 0$  se definirá a continuación. Decimos que  $X$  tiene una distribución Poisson con parámetro  $\lambda > 0$ , y escribimos  $X \sim \text{Poisson}(\lambda)$  cuando

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

Puede demostrarse que la función  $f(x)$  arriba definida es efectivamente una función de probabilidad para cada valor de  $\lambda > 0$  fijo, y para ello conviene recordar el desarrollo de la serie de Taylor de la función  $e^x$  alrededor del cero, es decir,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

La forma de esta función de probabilidad se muestra en la Figura 3.8 cuando  $\lambda = 2$ .

$x$	$f(x)$
0	0.1353353
1	0.2706706
2	0.2706706
3	0.180447
4	0.09022352
5	0.03608941
6	0.0120298
7	0.003437087
...	...

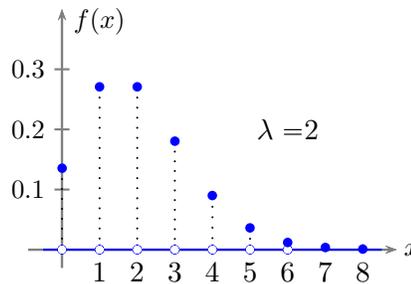


Figura 3.8

En R pueden obtenerse los valores de  $f(x)$  usando el siguiente comando

```
# dpois(x,lambda) evalúa f(x) de la distribución Poisson(lambda)
> dpois(3,2)
[1] 0.1804470
```

La función de distribución  $F(x)$  para esta distribución, como suma de los valores de  $f(x)$ , no tiene una expresión reducida y no la escribiremos, sin

embargo sus valores pueden encontrarse con facilidad en R mediante el siguiente comando:

```
# ppois(x,λ) evalúa  $F(x)$  de la distribución Poisson( $\lambda$ )
> ppois(3,2)
[1] 0.8571235
```

Después de algunos cálculos sencillos puede comprobarse que para esta distribución,

$$E(X) = \lambda,$$
$$\text{Var}(X) = \lambda.$$

**Simulación 3.7** Mediante el siguiente comando en R pueden generarse valores al azar de la distribución Poisson.

```
# rpois(k,λ) genera  $k$  valores al azar de la dist. Poisson( $\lambda$ )
> rpois(25,2)
[1] 0 3 3 3 1 1 3 0 0 2 0 3 1 1 0 3 5 4 1 0 1 2 1 2 1
```

**Ejemplo 3.12** En promedio se reciben 2 peticiones de acceso a una página web durante un minuto cualquiera. Utilice el modelo Poisson para calcular la probabilidad de que en un minuto dado:

- nadie solicite acceso a la página.
- se reciban mas de dos peticiones.

Solución: sea  $X$  el número de peticiones por minuto y supongamos que  $X$  tiene distribución Poisson( $\lambda$ ) con  $\lambda = 2$ . Para el primer inciso se tiene que

$$P(X = 0) = e^{-2} \frac{2^0}{0!} = 0.135.$$

Para el segundo inciso,

$$\begin{aligned}
 P(X > 2) = 1 - P(X \leq 2) &= 1 - (P(X = 0) + P(X = 1) + P(X = 2)) \\
 &= 1 - e^{-2} \left( \frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} \right) \\
 &= 0.323.
 \end{aligned}$$

•

Puede además demostrarse que cuando  $X \sim \text{bin}(n, p)$  y hacemos tender  $n$  a infinito y  $p$  a cero de tal forma que el producto  $np$  se mantenga constante igual a  $\lambda$ , entonces la variable aleatoria  $X$  adquiere la distribución Poisson con parámetro  $\lambda$ . Este resultado sugiere que cuando  $n$  es grande, la distribución binomial puede ser aproximada mediante la distribución Poisson de parámetro  $\lambda = np$ . Esto es particularmente útil pues el cálculo de probabilidades de la distribución binomial involucra el cálculo de factoriales y ello puede ser computacionalmente difícil para números grandes. El siguiente ejemplo ilustrará esta situación.

**Ejemplo 3.13** En promedio uno de cada 100 focos producidos por una máquina es defectuoso. Use la distribución Poisson para estimar la probabilidad de encontrar 5 focos defectuosos en un lote de 1000 focos.

*Solución:* sea  $X$  el número de focos defectuosos en el lote de 1000 focos. Entonces  $X$  tiene distribución  $\text{bin}(n, p)$  con  $n = 1000$  y  $p = 1/100$ . Nos piden calcular

$$P(X = 5) = \binom{1000}{5} (1/100)^5 (99/100)^{995} = 0.0374531116.$$

Usando la aproximación Poisson, con  $\lambda = np = 1000/100 = 10$ ,

$$P(X = 5) \approx e^{-10} \frac{10^5}{5!} = 0.0379841747.$$

•

Hemos definido a la variable aleatoria Poisson como el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado, supongamos

de longitud unitaria,  $[0, 1]$ . Suponga ahora que nos interesa observar las ocurrencias del evento en un intervalo de longitud diferente, por ejemplo  $[0, t]$ , con  $t > 0$ . Tal conteo de ocurrencias también sigue una distribución Poisson pero esta vez de parámetro  $\lambda t$ . Por ejemplo, si  $t = 2$ , entonces el número de ocurrencias del evento en el intervalo  $[0, 2]$  tiene distribución Poisson( $2\lambda$ ). El siguiente ejemplo ilustra esta situación.

**Ejemplo 3.14** El número de aviones que llegan a un aeropuerto internacional se considera como una cantidad aleatoria y se modela mediante una variable aleatoria con distribución Poisson con una frecuencia de 3 aviones cada 10 minutos. Es decir, la unidad de medición del tiempo son diez minutos. Entonces

- a) La probabilidad de que no llegue ningún avión en un periodo de 20 minutos (dos unidades de tiempo) es  $P(X = 0)$  con  $\lambda = 3(2) = 6$ .
- b) La probabilidad de que llegue sólo un avión en el minuto siguiente es  $P(X = 1)$  con  $\lambda = 3(1/10) = 3/10$ .
- c) La probabilidad de que lleguen dos o más aviones en un periodo de 15 minutos es  $P(X \geq 2)$  con  $\lambda = 3(1.5) = 4.5$ .

•

La distribución Poisson tiene algunas propiedades que resultan muy útiles en su aplicación. La siguiente es una de ellas.

**Proposición 3.2** Sean  $X_1$  y  $X_2$  dos variables aleatorias independientes con distribución Poisson( $\lambda_1$ ) y Poisson( $\lambda_2$ ), respectivamente. Entonces

$$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

**Demostración.** Para cualquier entero  $x \geq 0$ ,

$$\begin{aligned}
 P(X_1 + X_2 = x) &= \sum_{u=0}^x P(X_1 = u, X_2 = x - u) \\
 &= \sum_{u=0}^x P(X_1 = u) P(X_2 = x - u) \\
 &= \sum_{u=0}^x e^{-\lambda_1} \frac{\lambda_1^u}{u!} e^{-\lambda_2} \frac{\lambda_2^{x-u}}{(x-u)!} \\
 &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{x!} \sum_{u=0}^x \binom{x}{u} \lambda_1^u \lambda_2^{x-u} \\
 &= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^x}{x!}.
 \end{aligned}$$

•

El resultado anterior había sido demostrado antes en el Ejemplo 2.27 de la página 158 usando la propiedad de caracterización única de la f.g.p.

## Ejercicios

272. Demuestre que la función de probabilidad de la distribución Poisson( $\lambda$ ) es efectivamente una función de probabilidad.

Sugerencia:  $e^x = \sum_{k=0}^{\infty} x^k/k!$

273. Sea  $X$  con distribución Poisson( $\lambda$ ). Demuestre que:

- a)  $E(X) = \lambda$ .
- b)  $\text{Var}(X) = \lambda$ .

Sugerencia:  $x^2 = x(x-1) + x$ .

274. Sea  $f(x)$  la función de probabilidad de la distribución Poisson( $\lambda$ ). Demuestre que:

- a) **Fórmula iterativa:** para  $x \geq 0$  entero,

$$f(x+1) = \frac{\lambda}{x+1} f(x).$$

- b)  $f(x)$  es creciente de  $x$  a  $x + 1$  para valores enteros de  $x$  en el intervalo  $[0, \lambda - 1]$ .
- c)  $f(x)$  es decreciente de  $x$  a  $x + 1$  para valores enteros de  $x$  en el intervalo  $[\lambda - 1, \infty)$ .
- d)  $f(x)$  tiene un máximo en  $x^*$  definido como el entero más pequeño mayor o igual a  $\lambda - 1$ , es decir,  $x^* = \lceil \lambda - 1 \rceil$ .
- e)  $f(x^*) = f(x^* + 1)$  cuando  $x^*$  coincide con  $\lambda - 1$ . Por el inciso anterior, esto significa que  $f(x)$  tiene dos modas o máximos, uno en  $x^*$  y otro en  $x^* + 1$ .

275. **Aproximación a la distribución binomial.** Sea  $X$  una v.a. con distribución  $\text{bin}(n, p)$  tal que  $p = \lambda/n$  para  $n$  suficientemente grande, con  $\lambda > 0$  una constante. Demuestre que para cada  $k = 0, 1, \dots$

$$\lim_{n \rightarrow \infty} P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

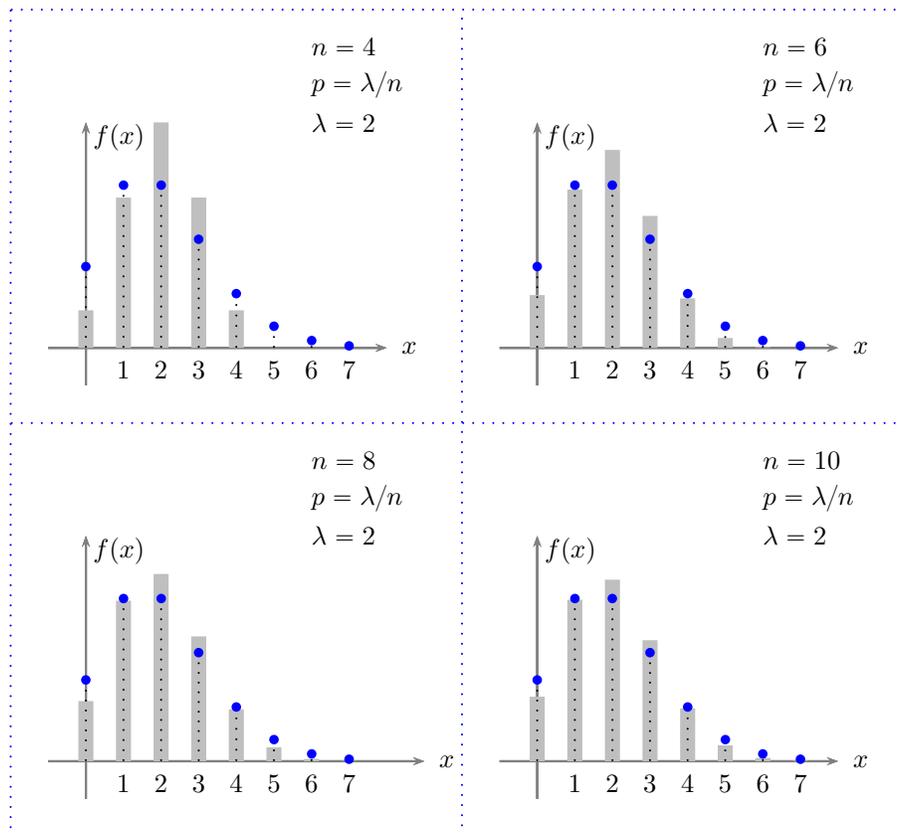
Esto quiere decir que la función de probabilidad binomial converge a la función de probabilidad Poisson, o en términos de aproximaciones, lo anterior sugiere que las probabilidades de la distribución binomial pueden aproximarse mediante las probabilidades Poisson cuando  $n$  es grande y  $p$  es pequeño de la forma  $\lambda/n$ . En la gráfica 3.9 se comparan las funciones de probabilidad de las distribuciones binomial( $n, p$ ) y Poisson( $\lambda$ ) con  $\lambda = 2$ ,  $p = \lambda/n$  para  $n = 4, 6, 8, 10$ . Se observa que conforme el parámetro  $n$  crece, las dos funciones son cada vez más parecidas. La gráfica de barras corresponde a la distribución binomial y gráfica de puntos corresponde a la distribución Poisson.

Sugerencia: Use la fórmula de Stirling

$$n! \approx \sqrt{2\pi n} n^{n+1/2} e^{-n}.$$

276. Sea  $X$  con distribución Poisson( $\lambda$ ). Demuestre que la función generadora de probabilidad de  $X$  es la función  $G(t)$  que aparece abajo. Usando esta función y sus propiedades demuestre nuevamente que  $E(X) = \lambda$  y  $\text{Var}(X) = \lambda$ .

$$G(t) = e^{\lambda(t-1)}.$$

Figura 3.9: Binomial *vs.* Poisson.

277. Sea  $X$  con distribución  $\text{Poisson}(\lambda)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo. Usando esta función y sus propiedades demuestre nuevamente que  $E(X) = \lambda$  y  $\text{Var}(X) = \lambda$ .

$$M(t) = e^{\lambda(e^t - 1)}.$$

278. El siguiente resultado fue demostrado en la Proposición 3.2 en la página 197. Sean  $X_1$  y  $X_2$  dos variables aleatorias independientes con distribución  $\text{Poisson}(\lambda_1)$  y  $\text{Poisson}(\lambda_2)$ , respectivamente. Use la función generadora de momentos para demostrar nuevamente que

$$X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2).$$

279. Sea  $X_1, X_2, \dots$  una sucesión de v.a.s independientes con distribución  $\text{Ber}(p)$  e independientes de  $N$  con distribución  $\text{Poisson}(\lambda)$ . Demuestre que la variable  $X$  definida a continuación tiene distribución  $\text{Poisson}(\lambda p)$ . Cuando  $N = 0$  la suma es vacía y se define como cero.

$$X = \sum_{i=1}^N X_i.$$

Sugerencia: Condicione sobre  $N$ .

280. Sea  $X$  una v.a. con distribución  $\text{Poisson}(\lambda)$  tal que  $\lambda \in (0, 1)$ . Demuestre que

$$E(X!) = \frac{1}{1 - \lambda} e^{-\lambda}.$$

281. En un libro de muchas páginas, el número de errores por página se modela mediante una v.a. con distribución  $\text{Poisson}$  con media uno. Encuentre la probabilidad de que una página seleccionada al azar contenga:

- a) ningún error.
- b) exactamente dos errores.
- c) al menos tres errores.

282. El número de semillas en una variedad de naranjas sigue una distribución Poisson de media 3. Calcule la probabilidad de que una naranja seleccionada al azar contenga:

- a) ninguna semilla.
- b) al menos dos semillas.
- c) a lo sumo tres semillas.

Con esto concluimos la revisión de algunas distribuciones de probabilidad de tipo discreto. Ahora estudiaremos algunas distribuciones de probabilidad de tipo continuo. No construiremos estas distribuciones a partir de experimentos aleatorios particulares como en el caso de algunas de las distribuciones discretas, mas bien, las definiremos sin mayor justificación.

Recordamos aquí nuevamente al lector que los parámetros de las distintas distribuciones de probabilidad pueden no ser usados de la misma forma en las distintas fuentes bibliográficas y paquetes computacionales.

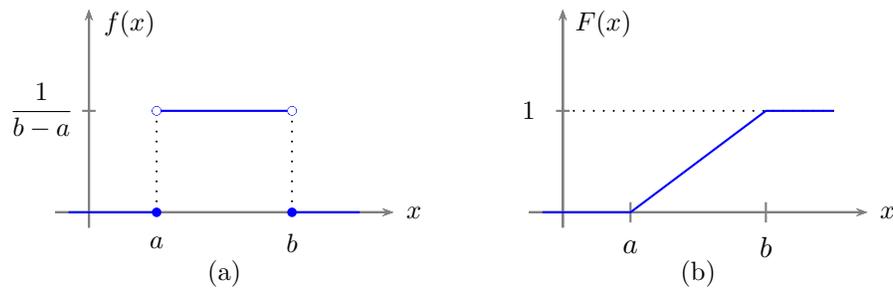
### 3.8. Distribución uniforme continua

Decimos que una variable aleatoria  $X$  tiene una distribución uniforme continua en el intervalo  $(a, b)$ , y escribimos  $X \sim \text{unif}(a, b)$ , cuando su función de densidad es

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

Esta distribución tiene como parámetros los números  $a$  y  $b$ . La gráfica general de esta función se muestra en la Figura 3.10(a), y es evidente que se trata de una función de densidad pues es no negativa e integra uno. Aunque es una función muy sencilla, sus valores pueden calcularse en R de la siguiente forma:

```
# dunif(x,a,b) evalúa f(x) de la distribución unif(a,b)
> dunif(2,-1,3)
[1] 0.25
```

Figura 3.10: Distribución uniforme( $a, b$ ).

Integrando esta función de densidad desde menos infinito hasta un punto  $x$  cualquiera, puede encontrarse la función de distribución, la cual tiene la siguiente expresión y cuya gráfica se muestra en la Figura 3.10(b).

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

Los valores de esta función pueden ser calculados en R usando el siguiente comando:

```
# punif(x,a,b) evalúa F(x) de la distribución unif(a,b)
> punif(2,-1,3)
[1] 0.75
```

Por otro lado es fácil verificar que

$$E(X) = (a + b)/2,$$

y

$$\text{Var}(X) = (b - a)^2/12.$$

Observe que la esperanza corresponde al punto medio del intervalo  $(a, b)$ . Además la varianza o dispersión crece cuando  $a$  y  $b$  se alejan uno del otro, y por el contrario, cuando estos parámetros están muy cercanos, la varianza es pequeña. Esta distribución es una de las más sencillas y sea usa naturalmente para cuando no se conoce mayor información de la variable aleatoria de interés, excepto que toma valores continuos dentro de cierto intervalo.

**Ejemplo 3.15** En el experimento aleatorio teórico de generar un número al azar  $X$  en un intervalo  $(a, b)$  se considera regularmente que  $X$  tiene distribución uniforme en dicho intervalo. Por ejemplo, algunos problemas estudiados antes sobre probabilidad geométrica hacen uso de esta distribución. •

**Simulación 3.8** Pueden generarse valores al azar en  $\mathbb{R}$  de la distribución uniforme continua usando el comando que aparece en el siguiente recuadro. Asigne valores de su preferencia a los parámetros  $a$  y  $b$ , y genere valores al azar de esta distribución.

```
# runif(k,a,b) genera k valores al azar de la dist. unif(a,b)
> runif(5,-1,3)
[1] 1.5889966 -0.1420308 2.5391841 0.4416415 0.7294366
```

## Ejercicios

283. Sea  $X$  con distribución  $\text{unif}(a, b)$ . Demuestre que:

- a)  $E(X) = \frac{a+b}{2}$ .  
 b)  $E(X^2) = \frac{1}{3}(a^2 + ab + b^2)$ .  
 c)  $\text{Var}(X) = \frac{(b-a)^2}{12}$ .

284. Sea  $X$  con distribución  $\text{unif}(-1, 1)$ . Demuestre que el  $n$ -ésimo momento de  $X$  está dado por

$$E(X^n) = \begin{cases} 0 & \text{si } n \text{ es impar,} \\ \frac{1}{n+1} & \text{si } n \text{ es par.} \end{cases}$$

285. Sea  $X$  con distribución  $\text{unif}(a, b)$ . Demuestre que el  $n$ -ésimo momento de  $X$  está dado por

$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

286. **Cuantil.** Sea  $p \in (0, 1)$ . Encuentre una expresión para el cuantil al 100 $p$ % de la distribución  $\text{unif}(a, b)$ .

Solución:  $c_p = a + (b - a)p$ .

287. **f.g.m.** Sea  $X$  con distribución  $\text{unif}(a, b)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo. Usando esta función y sus propiedades demuestre nuevamente que  $E(X) = (a + b)/2$  y  $\text{Var}(X) = (b - a)^2/12$ .

$$M(t) = \frac{e^{bt} - e^{at}}{t(b - a)}.$$

288. **Simulación.** Sea  $U$  una variable aleatoria con distribución  $\text{unif}(0, 1)$ . Sean  $a < b$  dos constantes. Demuestre que la variable  $X = a + (b - a)U$  tiene distribución  $\text{unif}(a, b)$ . Este es un mecanismo para obtener valores al azar de la distribución  $\text{unif}(a, b)$  a partir de valores al azar de la distribución  $\text{unif}(0, 1)$ . Esta última distribución aparece implementada en varios sistemas de cómputo.

289. Sea  $\epsilon > 0$ . Se escoge un número al azar  $X$  dentro del intervalo  $(0, 1)$ . Encuentre la probabilidad de que:

a)  $|X - 1/2| > \epsilon$ .

b)  $(2X - 1/2)^2 \leq \epsilon$ .

290. Se escoge un número al azar  $X$  dentro del intervalo  $(-1, 1)$ . Encuentre y grafique la función de densidad y de distribución de la variable:

a)  $Y = X + 1$ .

c)  $Y = X^2$ .

b)  $Y = |X|$ .

d)  $Y = X^3$ .

291. Sea  $X$  una variable aleatoria arbitraria cuyos posibles valores están contenidos en el intervalo  $(a, b)$ . Demuestre que:

a)  $a \leq E(X) \leq b$ .

b)  $0 \leq \text{Var}(X) \leq \frac{1}{4}(b - a)^2$ .

292. Sean  $\alpha$  y  $\beta > 0$  dos números fijos. Sea  $X$  con distribución  $\text{unif}(a, b)$ . Encuentre el valor de los parámetros  $a$  y  $b$  de tal forma que  $E(X) = \alpha$  y  $\text{Var}(X) = \beta$ .
293. **Una forma de aproximar  $\pi$ .** Escriba un programa en R para generar varias parejas al azar  $(a, b)$  con  $a$  y  $b$  con distribución uniforme dentro del intervalo  $(0, 1)$ , independiente un valor del otro. Si ocurre que  $b \leq \sqrt{1 - a^2}$ , entonces el punto  $(a, b)$  pertenece a la región sombreada de la Figura 3.11 y se considera un éxito, en caso contrario se considera un fracaso. Así, al repetir varias veces este procedimiento el cociente del número de éxitos  $n_E$  entre el número de ensayos  $n$  será una aproximación del área sombreada. Siendo esta área sombreada una cuarta parte del círculo unitario, su área es  $\pi/4$ . Grafique la función  $n \mapsto 4n_E/n$  para  $n = 1, 2, \dots, 100$  y comente su comportamiento conforme  $n$  crece.

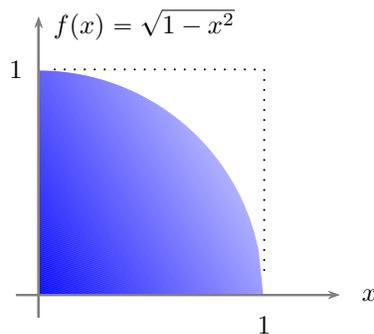


Figura 3.11

### 3.9. Distribución exponencial

Decimos que una variable aleatoria continua  $X$  tiene distribución exponencial con parámetro  $\lambda > 0$ , y escribimos  $X \sim \exp(\lambda)$ , cuando su función de densidad es

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

La gráfica de esta función cuando el parámetro  $\lambda$  toma el valor particular 3 se muestra en la Figura 3.12(a). La correspondiente función de distribución aparece a su derecha. Es muy sencillo verificar que la función  $f(x)$  arriba definida es efectivamente una función de densidad para cualquier valor del parámetro  $\lambda > 0$ . Se trata pues de una variable aleatoria continua con conjunto de valores el intervalo  $(0, \infty)$ . Esta distribución se usa para modelar tiempos de espera para la ocurrencia de un cierto evento. Los valores de  $f(x)$  pueden calcularse en R de la siguiente forma:

```
# dexp(x,lambda) evalúa f(x) de la distribución exp(lambda)
> dexp(0.5,3)
[1] 0.6693905
```

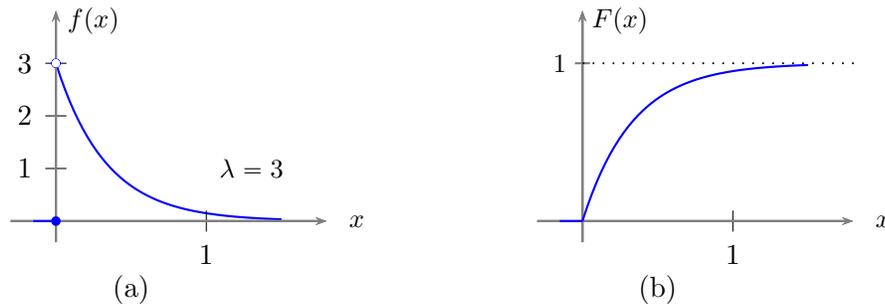


Figura 3.12

Integrando la función de densidad desde menos infinito hasta un valor arbitrario  $x$  se encuentra que la función de distribución tiene la expresión que aparece abajo y cuya gráfica se muestra en la Figura 3.12(b).

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

En R la función  $F(x)$  se calcula usando el siguiente comando:

```
# pexp(x,λ) evalúa F(x) de la distribución exp(λ)
> pexp(0.5,3)
[1] 0.7768698
```

Aplicando el método de integración por partes puede comprobarse que

$$E(X) = \frac{1}{\lambda},$$

$$\text{Var}(X) = \frac{1}{\lambda^2}.$$

**Simulación 3.9** Pueden generarse valores al azar en  $\mathbb{R}$  de la distribución exponencial usando el comando que aparece en el siguiente recuadro. Asigne un valor de su preferencia al parámetro  $\lambda$  y genere valores al azar de esta distribución.

```
# rexp(k,λ) genera k valores al azar de la dist. exp(λ)
> rexp(5,3)
[1] 0.53847926 0.19371105 0.32025823 0.07144621 0.20201383
```

**Ejemplo 3.16** Suponga que el tiempo en minutos que un usuario cualquiera permanece revisando su correo electrónico sigue una distribución exponencial de parámetro  $\lambda = 1/5$ . Calcule la probabilidad de que un usuario cualquiera permanezca conectado al servidor de correo

- menos de un minuto.
- mas de una hora.

Solución. Sea  $X$  el tiempo de conexión al servidor de correo. Para el primer inciso tenemos que

$$P(X < 1) = \int_0^1 1/5 e^{-x/5} dx = 0.181.$$

Para el segundo inciso,

$$P(X > 60) = \int_{60}^{\infty} 1/5 e^{-x/5} dx = 0.0000061.$$

## Ejercicios

294. Sea  $X$  con distribución  $\exp(\lambda)$ . Use la definición de esperanza y el método de integración por partes para demostrar que:

$$\begin{array}{ll} a) E(X) = \frac{1}{\lambda}. & c) E(X^3) = \frac{6}{\lambda^3}. \\ b) E(X^2) = \frac{2}{\lambda^2}. & d) \text{Var}(X) = \frac{1}{\lambda^2}. \end{array}$$

295. Sea  $X$  con distribución  $\exp(\lambda)$ . Use la fórmula (2.16) del Ejercicio 177 en la página 139 para encontrar la esperanza de las variables no negativas  $X$ ,  $X^2$  y  $X^3$  y demostrar nuevamente que:

$$\begin{array}{ll} a) E(X) = \frac{1}{\lambda}. & c) E(X^3) = \frac{6}{\lambda^3}. \\ b) E(X^2) = \frac{2}{\lambda^2}. & d) \text{Var}(X) = \frac{1}{\lambda^2}. \end{array}$$

296. **Momentos.** Sea  $X$  con distribución  $\exp(\lambda)$ . Demuestre que el  $n$ -ésimo momento de  $X$  es

$$E(X^n) = \frac{n!}{\lambda^n}.$$

297. Sea  $X$  con distribución  $\exp(\lambda)$  y sea  $c > 0$  una constante. Demuestre que

$$cX \sim \exp(\lambda/c).$$

298. Sea  $X$  con distribución  $\exp(\lambda)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo. Usando esta función y sus propiedades, demuestre nuevamente que  $E(X) = 1/\lambda$ ,  $E(X^2) = 2/\lambda^2$  y  $\text{Var}(X) = 1/\lambda^2$ .

$$M(t) = \frac{\lambda}{\lambda - t} \quad \text{para } t < \lambda.$$

299. Sea  $p \in (0, 1)$ . Encuentre el cuantil al  $100p\%$  de la distribución  $\exp(\lambda)$ . En particular, muestre que la mediana de esta distribución es  $(\ln 2)/\lambda$ .

$$\text{Solución: } c_p = -\frac{1}{\lambda} \ln(1 - p).$$

300. **Propiedad de pérdida de memoria.** Sea  $X$  con distribución exponencial de parámetro  $\lambda$ . Demuestre que para cualesquiera valores  $x, y \geq 0$ ,

$$P(X > x + y | X > y) = P(X > x).$$

301. **Discretización.** Sea  $X$  con distribución  $\exp(\lambda)$ . Demuestre que la variable aleatoria discreta definida a continuación tiene distribución  $\text{geo}(p)$  con  $p = 1 - e^{-\lambda}$ .

$$Y = \begin{cases} 0 & \text{si } 0 < X \leq 1, \\ 1 & \text{si } 1 < X \leq 2, \\ \dots & \dots \end{cases}$$

302. **Simulación: método de la función inversa.** Sea  $U$  una variable aleatoria con distribución  $\text{unif}(0, 1)$  y sea  $\lambda > 0$  una constante. Demuestre que la variable aleatoria  $X$  definida a continuación tiene distribución  $\exp(\lambda)$ .

$$X = -\frac{1}{\lambda} \ln(1 - U).$$

Este resultado permite obtener valores al azar de la distribución exponencial a partir de valores de la distribución uniforme continua.

303. **Coche en venta.** Un señor esta vendiendo su coche y decide aceptar la primera oferta que exceda \$50,000. Si las ofertas son variables aleatorias independientes con distribución exponencial de media \$45,000, encuentre:

- el número de ofertas recibidas hasta vender el coche.
- la probabilidad de que el precio de venta rebase \$55,000.
- el precio promedio de la venta del coche.

### 3.10. Distribución gama

La variable aleatoria continua  $X$  tiene una distribución gama con parámetros  $\alpha > 0$  y  $\lambda > 0$ , y escribimos  $X \sim \text{gama}(\alpha, \lambda)$ , si su función de densidad es

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

La gráfica de esta función de densidad para varios valores de los parámetros se muestra en la Figura 3.13. En la expresión anterior aparece el término  $\Gamma(\alpha)$ , el cual se conoce como la **función gama** y es de este hecho que la distribución adquiere su nombre. La función gama se define por medio de la siguiente integral:

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

para cualquier número real  $\alpha$  tal que esta integral sea convergente. Para evaluar la función gama es necesario substituir el valor de  $\alpha$  en el integrando y efectuar la integral infinita. En general, no necesitaremos evaluar esta integral para cualquier valor de  $\alpha$ , sólo para algunos pocos valores, principalmente enteros, y nos ayudaremos de las siguientes propiedades que no son difíciles de verificar:

- a)  $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ .
- b)  $\Gamma(\alpha + 1) = \alpha!$  si  $\alpha$  es un entero positivo.
- c)  $\Gamma(2) = \Gamma(1) = 1$ .
- d)  $\Gamma(1/2) = \sqrt{\pi}$ .

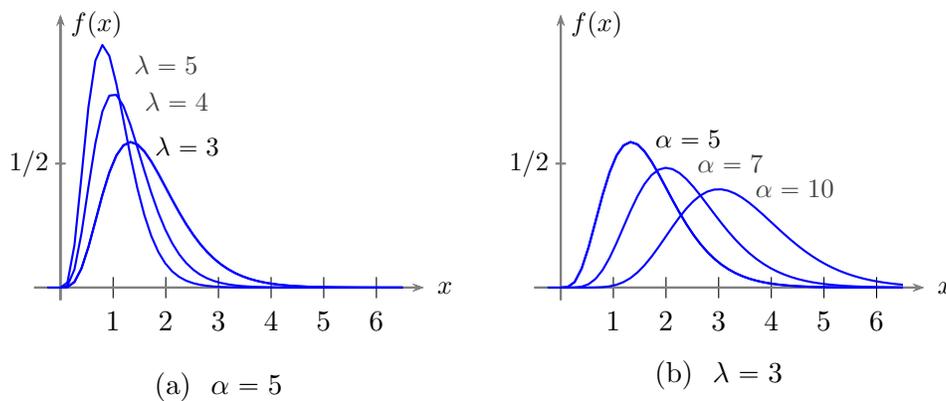


Figura 3.13

Así, a la función gama se le puede considerar como una generalización del factorial pues coincide con éste cuando el argumento es un número entero

positivo. En R pueden obtenerse los valores de la función  $\Gamma(x)$  mediante el comando `gamma(x)`. Los valores de la función de densidad  $f(x)$  se obtienen de la siguientes forma:

```
# dgamma(x,shape=alpha,rate=lambda) evalúa f(x) de la distribución
# gama(alpha, lambda)
> dgamma(2.5,shape=7,rate=3)
[1] 0.4101547
```

Observemos que la distribución exponencial es un caso particular de la distribución gama pues si en ésta se toma el parámetro  $\alpha$  igual a 1, se obtiene la distribución exponencial de parámetro  $\lambda$ .

Por otro lado, la función de distribución gama  $F(x)$  no tiene en general una expresión compacta, pero pueden calcularse con facilidad sus valores en R mediante el comando que aparece en el siguiente recuadro. Véase el Ejercicio 311 para conocer una fórmula para  $F(x)$  en un caso particular de sus parámetros.

```
# pgamma(x,shape=alpha,rate=lambda) evalúa F(x) de la distribución
# gama(alpha, lambda)
> pgamma(2.5,shape=7,rate=3)
[1] 0.6218453
```

Resolviendo un par de integrales se puede demostrar que

$$E(X) = \frac{\alpha}{\lambda},$$
$$\text{y } \text{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Cuando el parámetro  $\alpha$  es un número natural  $n$ , la distribución  $\text{gama}(n, \lambda)$  adquiere también el nombre de distribución Erlang( $n, \lambda$ ) y esta distribución puede obtenerse del siguiente resultado que es una aplicación inmediata de la f.g.m. y sus propiedades.

**Proposición 3.3** Sean  $X_1, \dots, X_n$  variables aleatorias independientes cada una de ellas con distribución  $\exp(\lambda)$ . Entonces

$$X_1 + \dots + X_n \sim \text{gama}(n, \lambda).$$

Así, una variable aleatoria con distribución  $\text{gama}(n, \lambda)$ , cuando  $n$  es un entero positivo, puede interpretarse como el tiempo acumulado de  $n$  tiempos de espera exponenciales independientes uno seguido del otro. Este resultado también indica un mecanismo para generar un valor al azar de la distribución  $\text{gama}(n, \lambda)$  a partir de  $n$  valores al azar de la distribución  $\exp(\lambda)$ .

**Simulación 3.10** Pueden generarse valores al azar en R de la distribución  $\text{gama}$  usando el comando que aparece en el siguiente recuadro. Asigne un valor de su preferencia a los parámetros  $\alpha$  y  $\lambda$  y genere valores al azar de esta distribución.

```
# rgamma(k,shape=α,rate=λ) genera k valores al azar de la
# distribución gama(α, λ)
> rgamma(5,shape=7,rate=3)
[1] 3.170814 1.433144 2.103220 1.662244 3.025049
```

## Ejercicios

304. Use la definición de la función  $\text{gama}$  para demostrar que la función de densidad  $\text{gama}$  efectivamente lo es.
305. Sea  $X$  con distribución  $\text{gama}(\alpha, \lambda)$ . Use la definición de esperanza para demostrar que:

$$\begin{aligned} a) \quad E(X) &= \frac{\alpha}{\lambda}. \\ b) \quad E(X^2) &= \frac{\alpha(\alpha + 1)}{\lambda^2}. \end{aligned}$$

$$c) \operatorname{Var}(X) = \frac{\alpha}{\lambda^2}.$$

306. Sea  $X$  con distribución  $\text{gama}(\alpha, \lambda)$ . Demuestre que el  $n$ -ésimo momento de  $X$  es

$$E(X^n) = \frac{\alpha(\alpha + 1) \cdots (\alpha + n - 1)}{\lambda^n}.$$

307. Sea  $X$  con distribución  $\text{gama}(\alpha, \lambda)$  y sea  $c > 0$  una constante. Demuestre que

$$cX \sim \text{gama}(\alpha, \lambda/c).$$

308. Sea  $X$  con distribución  $\text{gama}(\alpha, \lambda)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo. Usando esta función y sus propiedades demuestre nuevamente que  $E(X) = \alpha/\lambda$  y  $\operatorname{Var}(X) = \alpha/\lambda^2$ .

$$M(t) = \left( \frac{\lambda}{\lambda - t} \right)^\alpha \quad \text{para } t < \lambda.$$

309. Utilice la f.g.m. para demostrar que si  $X$  y  $Y$  son independientes con distribución  $\text{gama}(\alpha_1, \lambda)$  y  $\text{gama}(\alpha_2, \lambda)$  respectivamente, entonces

$$X + Y \sim \text{gama}(\alpha_1 + \alpha_2, \lambda).$$

310. Utilice la f.g.m. para demostrar la Proposición 3.3.

311. Sea  $X$  con distribución  $\text{gama}(n, \lambda)$  en donde  $n$  es un entero positivo. Use el resultado de la Proposición 3.3 para demostrar que la función de distribución de  $X$  adquiere la siguiente expresión: para cada  $x > 0$ ,

$$F(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x} = \sum_{k=n}^{\infty} \frac{(\lambda x)^k}{k!} e^{-\lambda x}.$$

312. Demuestre las siguientes propiedades de la función  $\text{gama}$ . Para el último inciso podría ayudar consultar la distribución normal que estudiaremos más adelante.

$$a) \Gamma(\alpha + 1) = \alpha \Gamma(\alpha).$$

$$b) \Gamma(\alpha + 1) = \alpha! \quad \text{si } \alpha \text{ es un entero positivo.}$$

$$c) \Gamma(2) = \Gamma(1) = 1.$$

$$d) \Gamma(1/2) = \sqrt{\pi}.$$

### 3.11. Distribución beta

Decimos que la variable aleatoria continua  $X$  tiene una distribución beta con parámetros  $a > 0$  y  $b > 0$ , y escribimos  $X \sim \text{beta}(a, b)$ , cuando su función de densidad es

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

El término  $B(a, b)$  se conoce como la **función beta**, y de allí adquiere su nombre esta distribución. La función beta se define como sigue:

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

para números reales  $a > 0$  y  $b > 0$ . Esta función está relacionada con la función gamma, antes mencionada, a través de la identidad:

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

Véase la sección de ejercicios para una lista de propiedades de esta función. La gráfica de la función de densidad beta se muestra en la Figura 3.14 para algunos valores de sus parámetros. En R pueden calcularse los valores de  $f(x)$  de la siguiente forma:

```
# dbeta(x,a,b) evalúa f(x) de la distribución beta(a,b)
> dbeta(0.3,1,2)
[1] 1.4
```

La correspondiente función de distribución no tiene una forma reducida y se escribe simplemente como sigue:

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \frac{1}{B(a, b)} \int_0^x u^{a-1} (1-u)^{b-1} du & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1, \end{cases}$$

y sus valores pueden calcularse en R usando el siguiente comando:

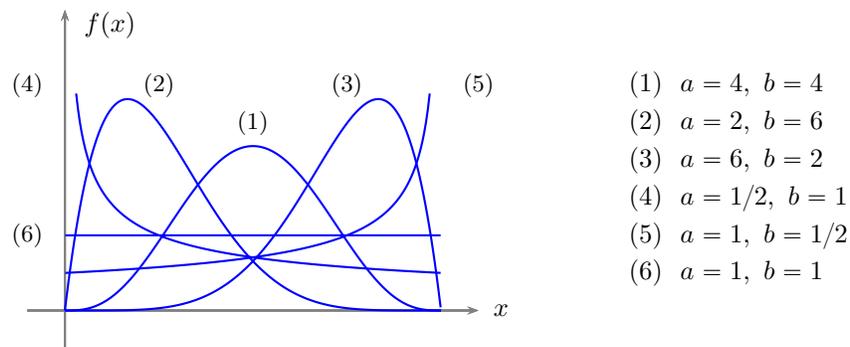


Figura 3.14

```
# pbeta(x,a,b) evalúa  $F(x)$  de la distribución  $\text{beta}(a,b)$ 
> pbeta(0.3,1,2)
[1] 0.51
```

Para la distribución  $\text{beta}(a, b)$ , usando una identidad que aparece en el Ejercicio 313, se puede demostrar sin mucha dificultad que

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

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

La función generadora de momentos no tiene una forma compacta para esta distribución. Por otro lado, en R se pueden generar valores al azar de la distribución beta de manera análoga a las otras distribuciones, esto es,

```
# rbeta(k,a,b) genera  $k$  valores al azar de la dist.  $\text{beta}(a,b)$ 
> rbeta(5,1,2)
[1] 0.18713260 0.07264413 0.08796477 0.15438134 0.29011107
```

La distribución beta puede obtenerse a partir de la distribución gamma como indica el siguiente resultado cuya demostración omitiremos.

**Proposición 3.4** Sean  $X$  y  $Y$  dos variables aleatorias independientes con distribución  $\text{gama}(a, \lambda)$  y  $\text{gama}(b, \lambda)$  respectivamente. Entonces

$$\frac{X}{X+Y} \sim \text{beta}(a, b).$$

Finalmente observamos que la distribución  $\text{beta}(a, b)$  se reduce a la distribución  $\text{unif}(0, 1)$  cuando  $a = b = 1$ .

## Ejercicios

313. Demuestre las siguientes propiedades de la función beta.

- a)  $B(a, b) = B(b, a)$ .
- b)  $B(a, 1) = 1/a$ .
- c)  $B(1, b) = 1/b$ .
- d)  $B(a+1, b) = \frac{a}{b} B(a, b+1)$ .
- e)  $B(a+1, b) = \frac{a}{a+b} B(a, b)$ .
- f)  $B(a, b+1) = \frac{b}{a+b} B(a, b)$ .
- g)  $B(1/2, 1/2) = \pi$ .

314. Demuestre que si  $X \sim \text{beta}(a, b)$ , entonces

$$1 - X \sim \text{beta}(b, a).$$

315. Sea  $X$  con distribución  $\text{beta}(a, b)$ . Demuestre que:

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

316. Sea  $X$  con distribución  $\text{beta}(a, b)$ . Demuestre que el  $n$ -ésimo momento de  $X$  es

$$E(X^n) = \frac{B(a+n, b)}{B(a, b)} = \frac{(a+n-1)(a+n-2)\cdots a}{(a+b+n-1)(a+b+n-2)\cdots(a+b)}.$$

317. Encuentre una expresión para la función de distribución  $F(x)$  de una variable aleatoria con distribución  $\text{beta}(a, b)$  para:

- a)  $a > 0, b = 1$ .
- b)  $a = 0, b > 1$ .

### 3.12. Distribución Weibull

La variable aleatoria continua  $X$  tiene una distribución Weibull con parámetros  $\alpha > 0$  y  $\lambda > 0$  si su función de densidad está dada por la siguiente expresión

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

A la constante  $\alpha$  se le llama parámetro de forma y a  $\lambda$  se le llama parámetro de escala. Se escribe  $X \sim \text{Weibull}(\alpha, \lambda)$ . La gráfica de la función de densidad para varios valores de sus parámetros se encuentra en la Figura 3.15 y su evaluación en R se obtiene usando el siguiente comando:

```
# dweibull(x, alpha, lambda) evalúa f(x) de la distribución Weibull(alpha, lambda)
> dweibull(2, 8, 2)
[1] 1.471518
```

Llevando a cabo un cambio de variable (véase el Ejercicio 319) puede demostrarse que la correspondiente función de distribución adquiere la siguiente forma simple:

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

cuyos valores pueden calcularse en R mediante el siguiente comando:

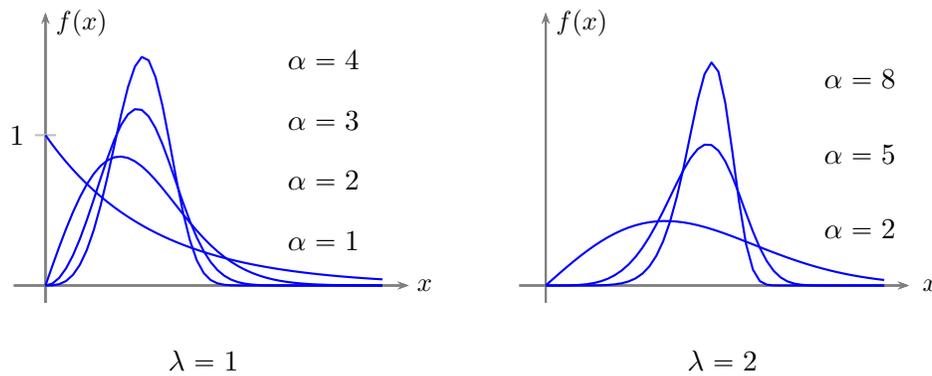


Figura 3.15

```
# pweibull(x,alpha,lambda) evalúa F(x) de la distribución Weibull(alpha,lambda)
> pweibull(2,8,2)
[1] 0.6321206
```

Aplicando la definición de la función gamma y después de algunos cálculos puede encontrarse que la esperanza y varianza de una variable aleatoria  $X$  con distribución Weibull( $\alpha, \lambda$ ) son

$$E(X) = \frac{1}{\lambda} \Gamma(1 + 1/\alpha),$$

$$\text{y } \text{Var}(X) = \frac{1}{\lambda^2} (\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)).$$

La distribución Weibull se ha utilizado en estudios de confiabilidad y durabilidad de componentes electrónicos y mecánicos. El valor de una variable aleatoria con esta distribución puede interpretarse como el tiempo de vida útil que tiene uno de estos componentes. Cuando el parámetro  $\alpha$  toma el valor uno, la distribución Weibull se reduce a la distribución exponencial de parámetro  $\lambda$ . En R se pueden generar valores al azar de la distribución Weibull usando el siguiente comando:

```
# rweibull(k,α,λ) genera k valores al azar de la distribución
# Weibull(α,λ)
> rweibull(5,8,2)
[1] 1.817331 1.768006 1.993703 1.915803 2.026141
```

## Ejercicios

318. Demuestre que la función de densidad Weibull efectivamente lo es.  
Sugerencia:  $\lambda u := (\lambda x)^\alpha$ .
319. Sea  $X$  con distribución Weibull( $\alpha, \lambda$ ). La correspondiente función de distribución es, para  $x > 0$ ,

$$F(x) = \int_0^x \lambda \alpha (\lambda y)^{\alpha-1} e^{-(\lambda y)^\alpha} dy.$$

Efectúe el cambio de variable  $\lambda u = (\lambda y)^\alpha$  en la integral para demostrar que

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

320. Sea  $X$  con distribución Weibull( $\alpha, \lambda$ ). Demuestre que:

- $E(X) = \frac{1}{\lambda} \Gamma(1 + 1/\alpha)$ .
- $E(X^2) = \frac{1}{\lambda^2} \Gamma(1 + 2/\alpha)$ .
- $\text{Var}(X) = \frac{1}{\lambda^2} (\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha))$ .

321. Sea  $X$  con distribución Weibull( $\alpha, \lambda$ ). Demuestre que el  $n$ -ésimo momento de  $X$  es

$$E(X^n) = \frac{1}{\lambda^n} \Gamma(1 + \frac{n}{\alpha}).$$

322. Invertiendo la función de distribución encuentre el cuantil al 100 $p$ % para una distribución Weibull( $\alpha, \lambda$ ).

323. **Simulación.** Sea  $U$  una variable aleatoria con distribución  $\text{unif}(0, 1)$  y sean  $\alpha > 0$  y  $\lambda > 0$  dos constantes. Demuestre que

$$X = \frac{1}{\lambda}(-\ln(1 - U))^{1/\alpha} \sim \text{Weibull}(\alpha, \lambda).$$

Este resultado permite obtener valores al azar de la distribución Weibull a partir de valores al azar de la distribución uniforme.

### 3.13. Distribución normal

Esta es posiblemente la distribución de probabilidad de mayor importancia. Decimos que la variable aleatoria continua  $X$  tiene una distribución normal si su función de densidad está dada por la siguiente expresión

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

en donde  $\mu \in \mathbb{R}$  y  $\sigma > 0$  son dos parámetros. Escribimos entonces  $X \sim N(\mu, \sigma^2)$ . La gráfica de esta función de densidad tiene forma de campana como se puede apreciar en la Figura 3.16, en donde se muestra además el significado geométrico de los dos parámetros.

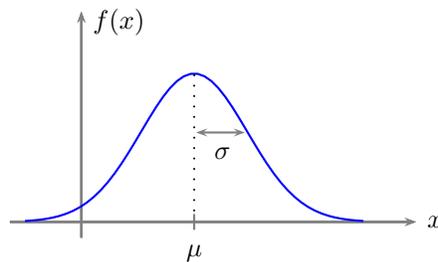


Figura 3.16

En R la evaluación de la función de densidad puede obtenerse usando el siguiente comando, aunque observe con cuidado que se usa la desviación estándar  $\sigma$  en su parametrización y no  $\sigma^2$ .

```
# dnorm(x,μ,σ) evalúa f(x) de la distribución N(μ,σ²)
> dnorm(1.5,3,2)
[1] 0.1505687
```

La correspondiente función de distribución es

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy,$$

pero resulta que esta integral es imposible de resolver y no puede encontrarse una expresión cerrada. Más adelante explicaremos la forma en que se han encontrado valores aproximados para esta integral. En R es muy sencillo obtener tales evaluaciones aproximadas utilizando el siguiente comando:

```
# pnorm(x,μ,σ) evalúa F(x) de la distribución N(μ,σ²)
> pnorm(1.5,3,2)
[1] 0.2266274
```

Por otro lado, usando integración por partes es posible demostrar que para una variable aleatoria  $X$  con distribución  $N(\mu, \sigma^2)$ ,

$$\begin{aligned} E(X) &= \mu, \\ \text{Var}(X) &= \sigma^2. \end{aligned}$$

Esto significa que la campana esta centrada en el valor del parámetro  $\mu$ , el cual puede ser negativo, positivo o cero, y que la campana se abre o se cierra de acuerdo a la magnitud del parámetro  $\sigma^2$ . El siguiente caso particular de la distribución normal es muy importante.

### Distribución normal estándar

Decimos que la variable aleatoria  $X$  tiene una distribución normal estándar si tiene una distribución normal con parámetros  $\mu = 0$  y  $\sigma^2 = 1$ . En este caso la función de densidad se reduce a la expresión

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

El resultado importante aquí es que siempre es posible transformar una variable aleatoria normal no estándar en una estándar mediante la siguiente operación cuya demostración se pide hacer en el Ejercicio 331.

**Proposición 3.5** Sea  $X$  con distribución  $N(\mu, \sigma^2)$ . Entonces

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1). \quad (3.7)$$

Al procedimiento anterior se le conoce con el nombre de **estandarización** y bajo tal transformación se dice que la variable  $X$  ha sido **estandarizada**. Este resultado parece muy modesto pero tiene una gran importancia operacional pues establece que el cálculo de las probabilidades de una variable aleatoria normal cualquiera se reduce al cálculo de las probabilidades para la normal estándar. Explicaremos ahora con más detalle esta situación. Suponga que  $X$  es una variable aleatoria con distribución  $N(\mu, \sigma^2)$  y que deseamos calcular, por ejemplo, la probabilidad de que  $X$  tome un valor en el intervalo  $(a, b)$ , es decir,  $P(a < X < b)$ . Tenemos entonces que

$$\begin{aligned} P(a < X < b) &= P(a - \mu < X - \mu < b - \mu) \\ &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) \\ &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right). \end{aligned}$$

Cada una de las igualdades anteriores es consecuencia de la igualdad de los eventos correspondientes. De esta forma una probabilidad que involucra a la variable  $X$  se ha reducido a una probabilidad que involucra a  $Z$ . De modo que únicamente necesitamos conocer las probabilidades de los eventos de  $Z$  para calcular las probabilidades de los eventos de la variable  $X$  que tiene parámetros arbitrarios. En términos de integrales el cálculo anterior es equivalente al siguiente en donde se lleva a cabo el cambio de variable

$y = (x - \mu)/\sigma$  en la integral,

$$\begin{aligned} P(a < X < b) &= \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right). \end{aligned}$$

Usaremos la letra  $Z$  para denotar a una variable aleatoria con distribución normal estándar.

### **Función de distribución $N(0, 1)$**

Es común denotar a la función de distribución de una variable aleatoria normal estándar como  $\Phi(x)$ , es decir,

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

cuyo significado geométrico se muestra en la Figura 3.17(a). Como hemos mencionado antes que no es posible resolver esta integral y para evaluarla se usan métodos numéricos para aproximar  $\Phi(x)$  para distintos valores de  $x$ . Aunque en R pueden encontrarse estos valores con el comando `pnorm(x, 0, 1)`, en la parte final del texto aparece una tabla con estos valores aproximados. Cada renglón de esta tabla corresponde a un valor de  $x$  hasta el primer dígito decimal, las distintas columnas corresponden al segundo dígito decimal. El valor que aparece en la tabla es  $\Phi(x)$ . Por ejemplo, el renglón marcado con 1.4 y la columna marcada con 0.05 corresponden al valor  $x = 1.45$ , tenemos entonces que  $\Phi(1.45) = 0.9265$ . Abajo aparecen algunos ejemplos que ilustran el uso de esta tabla. Observe además que para  $x \geq 3.5$ , la probabilidad  $\Phi(x)$  es muy cercana a uno, es decir, para esos valores de  $x$  la campana prácticamente ha decaído a cero en el lado derecho. Esto quiere decir que, con probabilidad cercana a uno, los valores que toma una variable aleatoria normal estándar están comprendidos entre  $-3.5$  y  $+3.5$ .

Por otro lado, a partir del hecho de que si  $X$  tiene distribución normal estándar, entonces la variable  $-X$  también tiene distribución normal estándar, puede demostrarse que

$$\Phi(-x) = 1 - \Phi(x).$$

Un argumento geométrico también puede utilizarse para darse cuenta de la validez de esta igualdad. En particular, este resultado ayuda a calcular valores de  $\Phi(x)$  para  $x$  negativos en tablas de la distribución normal como la presentada al final del texto en donde sólo aparecen valores positivos para  $x$ .

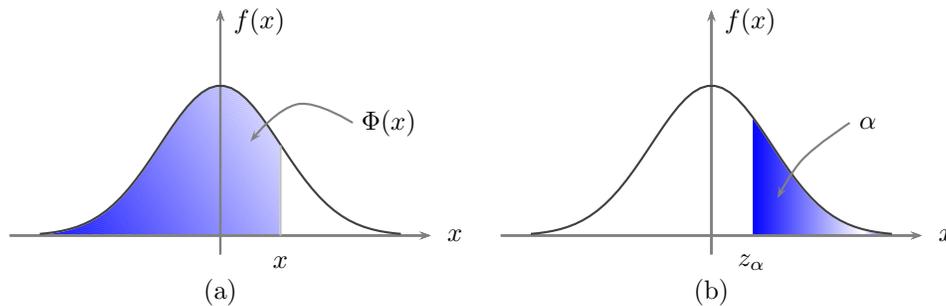


Figura 3.17

**Ejemplo 3.17** Use la tabla de la distribución normal estándar para comprobar que:

1.  $\Phi(1.65) = 0.9505$ .
2.  $\Phi(-1.65) = 0.0495$ .
3.  $\Phi(-1) = 0.1587$ .

•

**Ejemplo 3.18** Use la tabla de la distribución normal estándar para encontrar el valor de  $x$  tal que:

1.  $\Phi(x) = 0.3$ .
2.  $\Phi(x) = 0.75$ .

•

$$\text{Respuestas: } \begin{cases} 1. x = -0.53 \\ 2. x = 0.68 \end{cases}$$

**Ejemplo 3.19** Sea  $X$  con distribución  $N(5, 10)$ . Use el proceso de estandarización y la tabla de la distribución normal estándar para comprobar que:

1.  $P(X \leq 7) = 0.7357$ .
  2.  $P(0 < X < 5) = 0.2357$ .
  3.  $P(X > 10) = 0.0571$ .
- 

A continuación definiremos el número  $z_\alpha$ , el cual es usado con regularidad en las aplicaciones de la distribución normal.

**Notación**  $z_\alpha$ . Para cada valor de  $\alpha$  en el intervalo  $(0, 1)$ , el número  $z_\alpha$  denotará el cuantil al  $100(1 - \alpha)\%$  de la distribución normal estándar, es decir,

$$\Phi(z_\alpha) = 1 - \alpha.$$

El significado geométrico del número  $z_\alpha$  se muestra en la Figura 3.17(b).

**Ejemplo 3.20** Usando la tabla de la distribución normal estándar puede comprobarse que, de manera aproximada:

- a)  $z_{0.1} = 1.285$ .
  - b)  $z_{0.2} = 0.845$ .
-

Finalmente mencionaremos que en R se pueden generar valores al azar de la distribución normal haciendo uso del siguiente comando:

```
# rnorm(k,μ,σ) genera k valores al azar de la distribución
# N(μ,σ²)
> rnorm(5,3,2)
[1] 3.0408942 0.5529831 2.3426471 2.0050003 0.4448412
```

## Ejercicios

324. Demuestre que la función de densidad normal con parámetros  $\mu$  y  $\sigma^2$ :

- efectivamente es una función de densidad.
- tiene un máximo absoluto en  $x = \mu$ .
- tiene puntos de inflexión en  $x = \mu \pm \sigma$ .

325. Sea  $X$  con distribución  $N(\mu, \sigma^2)$ . Determine el valor de la constante  $c > 0$  tal que cumpla de manera aproximada la identidad que aparece abajo. Interprete este resultado.

$$P(\mu - c\sigma < X < \mu + c\sigma) = 0.99.$$

326. Demuestre que para cualquier  $x > 0$ ,

$$\frac{x}{1+x^2} e^{-x^2/2} \leq \int_x^\infty e^{-u^2/2} du \leq \frac{1}{x} e^{-x^2/2}.$$

327. Sea  $X$  con distribución  $N(\mu, \sigma^2)$ . Demuestre que:

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

328. Encuentre la moda y mediana de la distribución  $N(\mu, \sigma^2)$ .

329. **f.g.m.** Sea  $X$  con distribución  $N(\mu, \sigma^2)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo.



335. Sea  $X$  una variable aleatoria con distribución  $N(0, \sigma^2)$ . Calcule:

- a)  $E|X|$ .
- b)  $\text{Var}|X|$ .
- c)  $E(X^{2001})$ .

se puede modelar de manera aproximada como una variable aleatoria con distribución normal de media 30 minutos y desviación estándar 5 minutos. Cuánto tiempo debe asignarse a la aplicación del cuestionario si se desea que el 95% de las personas puedan completar el cuestionario?

- a)

### 3.14. Distribución ji-cuadrada

Decimos que la variable aleatoria continua  $X$  tiene una distribución ji-cuadrada con  $n$  grados de libertad ( $n > 0$ ), si su función de densidad está dada por la siguiente expresión:

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se trata de una variable aleatoria continua con posibles valores en el intervalo  $(0, \infty)$ . Esta distribución tiene un solo parámetro denotado aquí por la letra  $n$ , y al cual se le llama **grados de libertad**. A pesar de su aparente expresión complicada, no es difícil comprobar que  $f(x)$  es efectivamente una función de densidad, y para ello se utiliza la definición de la función gama. La gráfica de esta función de densidad para varios valores de su parámetro aparece en la Figura 3.18. Sus valores en R se encuentran de la forma siguiente:

```
# dchisq(x,n) evalúa f(x) de la distribución  $\chi^2(n)$ 
> dchisq(2,3)
[1] 0.2075537
```

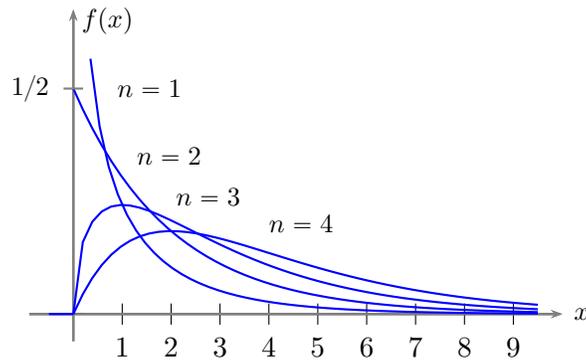


Figura 3.18

Escribiremos simplemente  $X \sim \chi^2(n)$ , en donde la letra griega  $\chi$  se pronuncia “ji” o también “chi”. Por lo tanto, la expresión “ $\chi^2(n)$ ” se lee “ji cuadrada con  $n$  grados de libertad”. Es interesante observar que la distribución  $\chi^2(n)$  se obtiene de la distribución  $\text{gama}(\alpha, \lambda)$  cuando  $\alpha = n/2$  y  $\lambda = 1/2$ . La expresión para la función de distribución no tiene una forma reducida: para  $x > 0$ ,

$$F(x) = \int_0^x \frac{1}{\Gamma(n/2)} \left(\frac{1}{2}\right)^{n/2} u^{n/2-1} e^{-u/2} du,$$

y sus valores pueden calcularse en R de la forma que aparece en el siguiente recuadro. Alternativamente en la parte final de este texto se encuentra una tabla con algunas de estas probabilidades.

```
# pchisq(x,n) evalúa F(x) de la distribución  $\chi^2(n)$ 
> pchisq(2,3)
[1] 0.4275933
```

A través de la reconstrucción de la distribución  $\chi^2$  con un parámetro diferente en la integral correspondiente, puede demostrarse sin mucha dificultad que

$$E(X) = n,$$

y

$$\text{Var}(X) = 2n.$$

La distribución ji-cuadrada puede obtenerse como indican los varios resultados que a continuación enunciaremos y cuyas demostraciones se piden desarrollar en la sección de ejercicios.

**Proposición 3.6** Si  $X \sim N(0, 1)$ , entonces

$$X^2 \sim \chi^2(1).$$

Es decir, el cuadrado de una variable aleatoria con distribución normal estándar tiene distribución ji-cuadrada con un grado de libertad. Vea el Ejercicio 337. Por otro lado, el siguiente resultado establece que la suma de dos variables aleatorias independientes con distribución ji-cuadrada tiene distribución nuevamente ji-cuadrada con grados de libertad la suma de los grados de libertad de los sumandos.

**Proposición 3.7** Sean  $X$  y  $Y$  variables aleatorias independientes con distribución  $\chi^2(n)$  y  $\chi^2(m)$  respectivamente, entonces

$$X + Y \sim \chi^2(n + m).$$

En el Ejercicio 343 se pide usar la f.g.m. para demostrar este resultado, el cual puede extenderse al caso cuando se tienen varias variables aleatorias independientes con distribución  $\chi^2$ . En particular, si  $X_1, \dots, X_n$  son variables independientes con distribución normal estándar, entonces la suma de los cuadrados  $X_1^2 + \dots + X_n^2$  tiene distribución  $\chi^2(n)$ . De este modo, si conocemos una forma de simular  $n$  valores al azar de la distribución normal estándar, la suma de los cuadrados de los números obtenidos será una observación de la distribución ji-cuadrada con  $n$  grados de libertad. Por último, mencionaremos el siguiente resultado el cual es utilizado en algunos procedimientos estadísticos.

**Proposición 3.8** Sean  $X_1, \dots, X_n$  variables aleatorias independientes cada una de ellas con distribución  $N(\mu, \sigma^2)$ . Entonces

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

en donde  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  y  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

A través del siguiente comando en R se pueden obtener valores al azar de la distribución  $\chi^2$ .

```
# rchisq(k,n) genera k valores al azar de la distribución
#  $\chi^2(n)$ 
> rchisq(5,3)
[1] 2.5946656 6.9019593 0.7172345 4.5362704 0.7995995
```

## Ejercicios

336. Compruebe que la función de densidad de la distribución  $\chi^2(n)$  efectivamente lo es.
337. Sea  $X$  una variable aleatoria continua. Demuestre que para  $x > 0$ ,

$$F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}).$$

Use este resultado para demostrar que si  $X \sim N(0, 1)$ , entonces  $X^2 \sim \chi^2(1)$ .

338. Sea  $X$  con distribución  $\chi^2(n)$ . Sea  $c > 0$  una constante y defina  $\alpha = n/2$  y  $\lambda = 2c$ . Demuestre que

$$cX \sim \text{gama}(\alpha, \lambda).$$

339. Compruebe que la distribución  $\text{gama}(\alpha, \lambda)$ , en donde  $\alpha = n/2$  con  $n \in \mathbb{N}$  y  $\lambda = 1/2$  se reduce a la distribución  $\chi^2(n)$ .

340. Sea  $X$  con distribución  $\chi^2(n)$ . Demuestre que:

- a)  $E(X) = n$ .
- b)  $E(X^2) = n(n + 2)$ .
- c)  $\text{Var}(X) = 2n$ .

341. Sea  $X$  con distribución  $\chi^2(n)$ . Demuestre que el  $m$ -ésimo momento de  $X$  es

$$E(X^m) = n(n + 2) \cdots (n + 2m - 2) = 2^m \frac{\Gamma(m + n/2)}{\Gamma(n/2)}.$$

342. Sea  $X$  con distribución  $\chi^2(n)$ . Demuestre que la función generadora de momentos de  $X$  es la función  $M(t)$  que aparece abajo. Usando esta función y sus propiedades demuestre nuevamente que  $E(X) = n$  y  $\text{Var}(X) = 2n$ .

$$M(t) = \left( \frac{1}{1 - 2t} \right)^{-n/2} \quad \text{para } t < 1/2.$$

343. Use la f.g.m. para demostrar que si  $X$  y  $Y$  son variables aleatorias independientes con distribución  $\chi^2(n)$  y  $\chi^2(m)$  respectivamente, entonces  $X + Y \sim \chi^2(n + m)$ .

344. Sea  $U$  con distribución unif(0, 1). Demuestre que

$$-2\ln(U) \sim \chi^2(2) = \exp(1/2).$$

### 3.15. Distribución $t$

Decimos que la variable aleatoria continua  $X$  tiene una distribución  $t$  con  $n > 0$  grados de libertad si su función de densidad está dada por la siguiente expresión

$$f(x) = \frac{\Gamma((n + 1)/2)}{\sqrt{n\pi} \Gamma(n/2)} (1 + x^2/n)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$

En tal caso se escribe  $X \sim t(n)$ , en donde  $n$  es un número real positivo aunque tomaremos principalmente el caso cuando  $n$  es entero positivo. La gráfica de esta función de densidad aparece en la Figura 3.19 y sus valores pueden calcularse en R con ayuda del siguiente comando.

```
# dt(x,n) evalúa f(x) de la distribución t(n)
> dt(1,3)
[1] 0.2067483
```

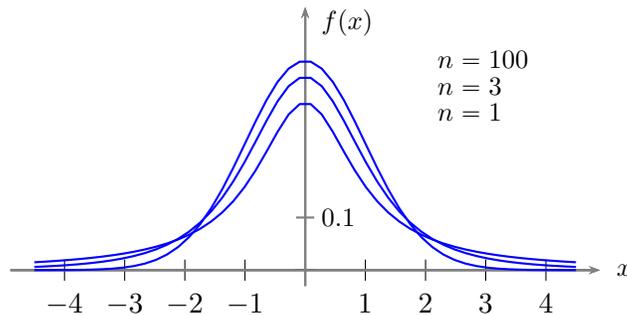


Figura 3.19

En la Figura anterior puede apreciarse el parecido de la función de densidad  $t(n)$  con la función de densidad normal estándar. En esta misma Figura está graficada también la función de densidad normal estándar pero ésta se empalma completamente con la función de densidad  $t(n)$  con  $n = 100$ . En el límite cuando  $n \rightarrow \infty$  ambas densidades coinciden, véase el Ejercicio 350. La función de distribución no tiene una expresión simple y la dejaremos indicada como la integral correspondiente, es decir,

$$F(x) = \int_{-\infty}^x \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \Gamma(n/2)} (1 + u^2/n)^{-(n+1)/2} du,$$

y cuyos valores pueden encontrarse en una tabla al final del texto o bien en R mediante el siguiente comando:

```
# pt(x,n) evalúa  $F(x)$  de la distribución  $t(n)$   
> pt(1,3)  
[1] 0.8044989
```

Llevando a cabo la integral correspondiente no es difícil demostrar que

$$E(X) = 0,$$
$$\text{Var}(X) = \frac{n}{n-2} \quad \text{para } n > 2.$$

La distribución  $t$  es un ejemplo de distribución para la cual no existe la función generadora de momentos. Esta distribución se puede encontrar cuando se estudian ciertas operaciones entre otras variables aleatorias. Por simplicidad en la exposición omitiremos la demostración de los siguientes resultados.

**Proposición 3.9** Si  $X \sim N(0, 1)$  y  $Y \sim \chi^2(n)$  son independientes, entonces

$$\frac{X}{\sqrt{Y/n}} \sim t(n).$$

En el estudio y aplicación de la estadística matemática, se necesitan realizar operaciones como la indicada en la proposición anterior. Por otro lado, este resultado sugiere un mecanismo para generar simulaciones de los valores que toma una variable aleatoria con distribución  $t(n)$ . Para ello se pueden generar  $n$  observaciones de la distribución normal estándar, y con ello conformar una observación de la distribución  $\chi^2(n)$  como fue explicado antes. Se necesita una observación adicional de la distribución normal estándar que será el valor de  $X$ , según la fórmula de la proposición anterior, se hace el cociente indicado y el resultado será un valor de la distribución  $t(n)$ . En el siguiente contexto también aparece la distribución  $t$ .

**Proposición 3.10** Sean  $X_1, \dots, X_n$  variables aleatorias independientes cada una de ellas con distribución  $N(\mu, \sigma^2)$ . Entonces

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

en donde  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  y  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

Por último mencionaremos que se pueden generar valores al azar de la distribución  $t$  en el paquete R usando el siguiente comando:

```
# rt(k,n) genera k valores al azar de la distribución t(n)
> rt(5,3)
[1] 0.06769745 -0.33693291 -0.36182444 1.68520735 -0.02326697
```

## Ejercicios

345. Demuestre que la función de densidad  $t(n)$  efectivamente lo es.

346. Sea  $X$  con distribución  $t(n)$ . Demuestre que:

- a)  $E(X) = 0$  para  $n > 1$ .
- b)  $\text{Var}(X) = \frac{n}{n-2}$  para  $n > 2$ .

347. Sea  $X$  con distribución  $t(n)$ . Demuestre que:

$$E(X^m) = \begin{cases} 0 & \text{si } m \text{ es impar y } 2 \leq m < n, \\ \frac{\Gamma((m+1)/2)\Gamma((n-m)/2)n^{m/2}}{\sqrt{\pi}\Gamma(n/2)} & \text{si } m \text{ es par y } 2 \leq m < n, \\ \text{"no existe"} & \text{si } m \geq n. \end{cases}$$

348. Encuentre la mediana y moda de la distribución  $t(n)$ .

349. Demuestre que no existe la f.g.m. para la distribución  $t(n)$ .

350. **Convergencia a la dist. normal estándar.** Sea  $f(x)$  la función de densidad  $t(n)$ . Demuestre que

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

### 3.16. Distribución $F$

La variable aleatoria continua  $X$  tiene una distribución  $F$  de Fisher con parámetros  $a > 0$  y  $b > 0$  si su función de densidad está dada por la siguiente expresión:

$$f(x) = \begin{cases} \frac{\Gamma(\frac{a+b}{2})}{\Gamma(\frac{a}{2})\Gamma(\frac{b}{2})} \left(\frac{a}{b}\right)^{a/2} x^{a/2-1} \left(1 + \frac{a}{b}x\right)^{-(a+b)/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

En este caso se escribe  $X \sim F(a, b)$ . Una gráfica de esta función de densidad aparece en la Figura 3.20, y sus valores pueden ser calculados en R usando el siguiente comando:

```
# df(x,n) evalúa f(x) de la distribución F(a,b)
> df(0.5,4,10)
[1] 0.669796
```

La función de distribución no tiene una expresión reducida y la indicaremos simplemente como la integral correspondiente, es decir, para  $x > 0$ ,

$$F(x) = \int_0^x \frac{\Gamma(\frac{a+b}{2})}{\Gamma(\frac{a}{2})\Gamma(\frac{b}{2})} \left(\frac{a}{b}\right)^{a/2} u^{a/2-1} \left(1 + \frac{a}{b}u\right)^{-(a+b)/2} du,$$

cuyos valores en R pueden encontrarse usando el siguiente comando:

```
# pf(x,n) evalúa F(x) de la distribución F(a,b)
> pf(0.5,4,10)
[1] 0.2632245
```

Para esta distribución se puede demostrar que

$$E(X) = \frac{b}{b-2}, \quad \text{si } b > 2,$$

y

$$\text{Var}(X) = \frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}, \quad \text{si } b > 4.$$

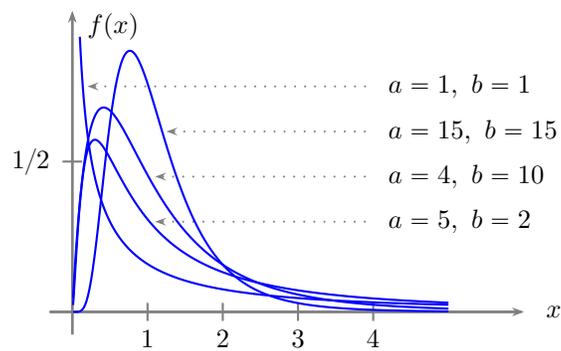


Figura 3.20

La distribución  $F$  aparece como resultado de la siguiente operación entre variables aleatorias.

**Proposición 3.11** Sean  $X \sim \chi^2(a)$  y  $Y \sim \chi^2(b)$  variables aleatorias independientes. Entonces

$$\frac{X/a}{Y/b} \sim F(a, b).$$

Por último, mencionaremos que se pueden generar valores al azar de la distribución  $F$  en el paquete R usando un comando similar a los anteriores:

```
# rf(k,a,b) genera k valores al azar de la distribución F(a,b)
> rf(5,4,10)
[1] 0.57341208 0.36602858 1.05682859 0.08009087 3.80035154
```

## Ejercicios

351. Demuestre que la función de densidad  $F(a, b)$  es efectivamente una función de densidad.

352. Sea  $X$  con distribución  $F(a, b)$ . Demuestre que:

$$a) E(X) = \frac{b}{b-2}, \quad \text{si } b > 2.$$

$$b) E(X^2) = \frac{b^2(a+2)}{a(b-2)(b-4)}, \quad \text{si } b > 4.$$

$$c) \text{Var}(X) = \frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}, \quad \text{si } b > 4.$$

353. Sea  $X$  con distribución  $F(a, b)$ . Demuestre que el  $n$ -ésimo momento de  $X$ , para  $2n < b$ , es

$$E(X^n) = \left(\frac{b}{a}\right)^n \frac{\Gamma(a/2 + n)}{\Gamma(a/2)} \frac{\Gamma(b/2 - n)}{\Gamma(b/2)}.$$

354. Demuestre que la f.g.m. de la distribución  $F(a, b)$  no existe.

355. Demuestre que si  $X \sim t(n)$  entonces

$$X^2 \sim F(1, n).$$

356. Demuestre que si  $X \sim F(a, b)$  entonces

$$\frac{1}{X} \sim F(b, a).$$

Con esto concluimos una revisión elemental de algunas distribuciones de probabilidad continuas. Recuerde el lector que existen muchas más distribuciones de este tipo, algunas más conocidas que otras, pero todas ellas útiles como modelos probabilísticos en las muy diversas áreas de aplicación de la probabilidad.



## Capítulo 4

# Vectores aleatorios

Este capítulo contiene una breve introducción al tema de variables aleatorias multidimensionales o también llamadas vectores aleatorios. Para hacer la escritura corta se consideran únicamente vectores aleatorios de dimensión dos, aunque las definiciones y resultados que se mencionan pueden extenderse fácilmente, en la mayoría de los casos, para vectores de dimensión superior. Para el material que se presenta a continuación sería provechoso contar con algunos conocimientos elementales del cálculo diferencial e integral en varias variables, o por lo menos mantener la calma cuando parezca que los símbolos matemáticos no tienen ningún sentido.

### 4.1. Vectores aleatorios

**Definición 4.1** Un **vector aleatorio** de dimensión dos es un vector de la forma  $(X, Y)$  en donde cada coordenada es una variable aleatoria. De manera análoga se definen vectores aleatorios multidimensionales  $(X_1, \dots, X_n)$ .

Se dice que un vector aleatorio es discreto, o continuo, si las todas las variables aleatorias que lo conforman lo son. Por simplicidad, considerare-

mos únicamente vectores aleatorios cuyas coordenadas son variables aleatorias todas discretas, o continuas, pero no combinaciones de ellas. Un vector aleatorio  $(X, Y)$  puede considerarse como una función de  $\Omega$  en  $\mathbb{R}^2$  como se muestra en la Figura 4.1.

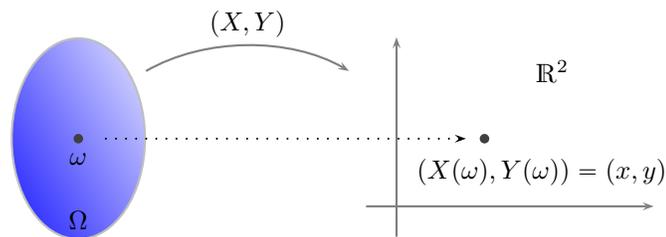


Figura 4.1: Vector aleatorio.

Es decir, el vector  $(X, Y)$  evaluado en  $\omega$  es  $(X, Y)(\omega) = (X(\omega), Y(\omega))$  con posible valor  $(x, y)$ . Nuevamente observe que el vector con letras mayúsculas  $(X, Y)$  es el vector aleatorio, mientras que el vector con letras minúsculas  $(x, y)$  es un punto en el plano. Así, el vector  $(X(\omega), Y(\omega))$  representa la respuesta conjunta de dos preguntas o mediciones efectuadas a un mismo elemento  $\omega$  del espacio muestral  $\Omega$ . A veces la información de la que se dispone acerca de un fenómeno está agrupada de esta forma. En nuestro caso hemos mencionado únicamente dos variables aleatorias pero vectores de dimensión mayor son posibles.

**Ejemplo 4.1** Suponga que tenemos una población de mujeres y que la variable  $X$  toma el valor 1 cuando la mujer es fumadora y cero cuando no lo es. Sea  $Y$  la variable que registra el número de hijos de una mujer dada. Entonces el vector  $(X, Y)$  puede tomar los valores:

$$(0, 0), (0, 1), (0, 2), \dots$$

$$(1, 0), (1, 1), (1, 2), \dots$$

Para una población particular de mujeres podríamos conformar una tabla con la frecuencia de cada una de estos valores del vector, por ejemplo,

$x/y$	0	1	2	3	4	5
0	2	5	8	3	2	0
1	10	9	2	0	1	1

**Ejemplo 4.2** Suponga que se cuenta con una población de personas que participan en un proceso de elección. Se escoge a uno de los votantes al azar y el vector  $(X(\omega), Y(\omega))$  puede representar el nivel económico y la preferencia electoral del votante  $\omega$ .

Estudiaremos a continuación algunas funciones asociadas a vectores aleatorios, las cuales son análogas al caso unidimensional estudiado antes.

## 4.2. Función de probabilidad conjunta

Estudiaremos primero el caso de vectores aleatorios discretos. La situación es muy similar al caso unidimensional.

**Definición 4.2** La **función de probabilidad** del vector aleatorio discreto  $(X, Y)$ , en donde  $X$  toma los valores  $x_1, x_2, \dots$  y  $Y$  toma los valores  $y_1, y_2, \dots$ , es la función  $f(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$  dada por

$$f(x, y) = \begin{cases} P(X = x, Y = y) & \text{si } (x, y) \in \{x_1, x_2, \dots\} \times \{y_1, y_2, \dots\}, \\ 0 & \text{en otro caso.} \end{cases}$$

Es decir, la función  $f(x, y)$  es la probabilidad de que la variable  $X$  tome el valor  $x$  y al mismo tiempo la variable  $Y$  tome el valor  $y$ . Tal función se llama también **función de probabilidad conjunta** de las variables  $X$  y  $Y$ , y para enfatizar este hecho a veces se escribe  $f_{X,Y}(x, y)$ , pero en general omitiremos los subíndices para hacer la notación más corta pero asociando el valor  $x$

a la variable  $X$  y el valor  $y$  a la variable  $Y$ . Haremos uso de los subíndices cuando sea necesario especificar las variables aleatorias en estudio.

Toda función  $f(x, y)$  de la forma anterior cumple las siguientes dos propiedades, e inversamente, toda función definida sobre  $\mathbb{R}^2$  que sea cero excepto en un conjunto discreto de parejas  $(x, y)$  que cumpla estas propiedades se llama **función de probabilidad bivariada** o **conjunta**, sin necesidad de contar con dos variables aleatorias previas que la definan.

a)  $f(x, y) \geq 0$ .

b)  $\sum_{x,y} f(x, y) = 1$ .

Otra forma equivalente de presentar a la función de probabilidad de un vector discreto  $(X, Y)$  es a través de una tabla como la siguiente

$x/y$	$y_1$	$y_2$	$\dots$
$x_1$	$f(x_1, y_1)$	$f(x_1, y_2)$	$\dots$
$x_2$	$f(x_2, y_1)$	$f(x_2, y_2)$	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$

También se pueden elaborar gráficas en  $\mathbb{R}^3$  de las funciones de probabilidad bivariadas y su aspecto general se muestra en la Figura 4.2. Observe que sería difícil graficar funciones de probabilidad de vectores aleatorios de dimensión 3 y superiores.

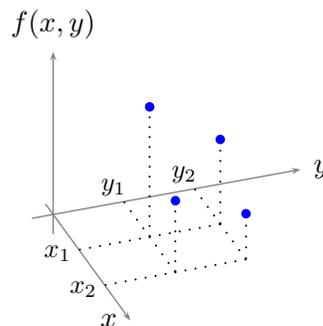


Figura 4.2

El cálculo de probabilidades de eventos relativos a un vector aleatorio discreto  $(X, Y)$  con función de probabilidad  $f(x, y)$  se lleva a cabo de la siguiente forma: si  $A$  y  $B$  son dos conjuntos de números reales, entonces la probabilidad del evento  $(X \in A) \cap (Y \in B)$  se calcula como sigue:

$$P(X \in A, Y \in B) = \sum_{x \in A} \sum_{y \in B} f(x, y).$$

**Ejemplo 4.3** Considere el vector aleatorio discreto  $(X, Y)$  con función de probabilidad dada por la siguiente tabla:

$x/y$	0	1
-1	0.3	0.1
1	0.4	0.2

De este arreglo se entiende que la variable  $X$  toma valores en el conjunto  $\{-1, 1\}$ , mientras que  $Y$  toma valores en  $\{0, 1\}$ . Además, las probabilidades conjuntas están dadas por las entradas de la tabla, por ejemplo,  $P(X = -1, Y = 0) = 0.3$ , esto es, la probabilidad de que  $X$  tome el valor  $-1$  y al mismo tiempo  $Y$  tome el valor  $0$  es  $0.3$ . La misma información puede escribirse de la siguiente manera:

$$f(x, y) = P(X = x, Y = y) = \begin{cases} 0.3 & \text{si } x = -1, y = 0, \\ 0.1 & \text{si } x = -1, y = 1, \\ 0.4 & \text{si } x = 1, y = 0, \\ 0.2 & \text{si } x = 1, y = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Como todos estos valores son probabilidades, naturalmente son no negativos y todos ellos suman uno. Por lo tanto,  $f(x, y)$  es efectivamente una función de probabilidad bivariada. •

**Ejemplo 4.4** Encontraremos la constante  $c$  que hace a la siguiente una función de probabilidad conjunta.

$$f(x, y) = \begin{cases} cxy & \text{si } (x, y) \in \{1, 2\} \times \{1, 2\}, \\ 0 & \text{otro caso.} \end{cases}$$

Los posibles valores del vector  $(X, Y)$  son  $(1, 1)$ ,  $(1, 2)$ ,  $(2, 1)$  y  $(2, 2)$ , con probabilidades respectivas  $c$ ,  $2c$ ,  $2c$  y  $4c$ . Como la suma de estas probabilidades debe ser uno, se llega a la ecuación  $9c = 1$ , de donde se obtiene que  $c = 1/9$ . •

Veamos ahora la situación en el caso de vectores aleatorios continuos.

**Definición 4.3** Sea  $(X, Y)$  un vector aleatorio continuo. Se dice que la función integrable y no negativa  $f(x, y) : \mathbb{R}^2 \rightarrow [0, \infty)$  es la **función de densidad** del vector  $(X, Y)$ , o bien que es la **función de densidad conjunta** de las variables  $X$  y  $Y$  si para todo par  $(x, y)$  en  $\mathbb{R}^2$  se cumple la igualdad

$$P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du. \quad (4.1)$$

La doble integral que aparece en (4.1) representa el volumen bajo la superficie dada por la función  $f(u, v)$  sobre la región que se encuentra a la izquierda y abajo del punto  $(x, y)$ . Toda función de densidad  $f(x, y)$  de estas características satisface las siguientes dos propiedades:

- a)  $f(x, y) \geq 0$ .
- b)  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$ .

Recíprocamente, decimos que una función  $f(x, y) : \mathbb{R}^2 \rightarrow [0, \infty)$  es una **función de densidad conjunta** o **bivariada** si cumple con las dos condiciones arriba mencionadas. El aspecto general de una función de densidad conjunta de dos variables aleatorias continuas es el de una superficie en  $\mathbb{R}^3$  como la que se muestra en la Figura 4.3.

El cálculo de probabilidades de eventos relativos a un vector aleatorio continuo  $(X, Y)$  con función de densidad  $f(x, y)$  se lleva a cabo de la siguiente forma: si  $a < b$  y  $c < d$ , entonces la probabilidad del evento  $(a < X < b) \cap (c < Y < d)$  se calcula como el volumen bajo la super-

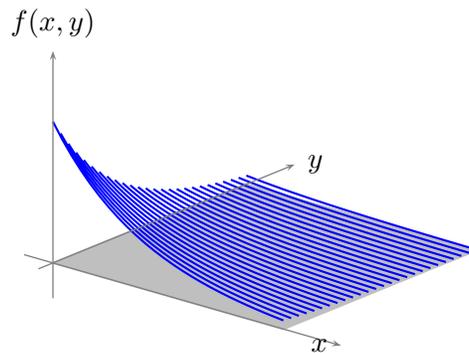


Figura 4.3

ficie  $f(x, y)$  en el rectángulo  $(a, b) \times (c, d)$ , es decir,

$$P(a < X < b, c < Y < d) = \int_a^b \int_c^d f(x, y) dy dx.$$

**Ejemplo 4.5** La siguiente función es una de las funciones de densidad conjunta más sencillas, se trata de la distribución uniforme continua bidimensional. Sean  $a < b$ ,  $c < d$ , y defina la función

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)} & \text{si } a < x < b, c < y < d, \\ 0 & \text{en otro caso,} \end{cases}$$

cuya gráfica aparece en la Figura 4.4. Se trata de una función constante en el rectángulo  $(a, b) \times (c, d)$ . Esta función es de densidad pues es no negativa e integra uno sobre  $\mathbb{R}^2$ . La doble integral sobre  $\mathbb{R}^2$  es simplemente el volumen del paralelepípedo que se muestra en la Figura 4.4. •

**Ejemplo 4.6** Comprobaremos que la siguiente función es de densidad.

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

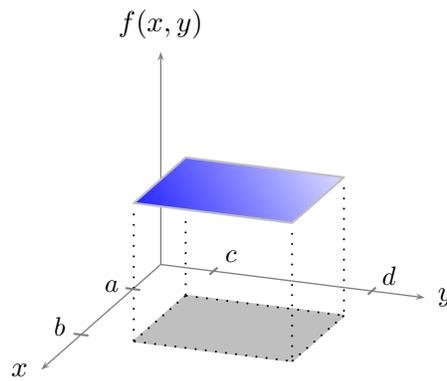


Figura 4.4

Claramente  $f(x, y) \geq 0$  para cualquier  $(x, y) \in \mathbb{R}^2$ . Resta verificar que la función integra uno sobre el plano. Se puede comprobar que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^1 \int_0^1 (x + y) dx dy = \frac{1}{2} + \frac{1}{2} = 1.$$

La gráfica de esta función se muestra en la Figura 4.5. •

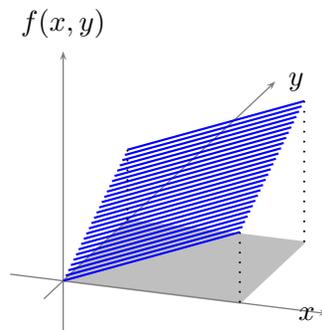


Figura 4.5

**Ejemplo 4.7** Encontraremos la constante  $c$  para que la siguiente función sea de densidad.

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

La constante  $c$  debe ser tal que la función  $f(x, y)$  sea no negativa y que su integral sobre todo el plano sea uno. De esta última condición obtenemos que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = \int_0^1 \int_0^y cxy \, dx \, dy = \int_0^1 \frac{c}{2} y^3 \, dy = \frac{c}{8}.$$

Por lo tanto,  $c = 8$ . La gráfica de la función  $f(x, y)$  se muestra en el Figura 4.6. •

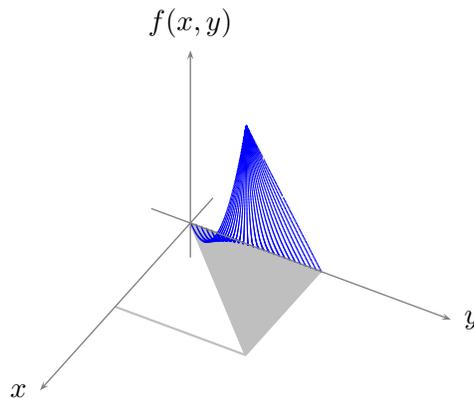


Figura 4.6

## Ejercicios

357. Sean  $X$  y  $Y$  dos variables aleatorias discretas con función de probabilidad conjunta dada por la siguiente tabla.

$x/y$	0	1	2
0	1/30	2/30	3/30
1	4/30	0	6/30
2	5/30	4/30	5/30

Encuentre:

- |                              |                                   |
|------------------------------|-----------------------------------|
| a) $P(X > 0, Y \geq 1)$ .    | f) $P(Y \leq 1   X = 1)$ .        |
| b) $P(X \leq 1, Y \geq 1)$ . | g) $P(XY = 0)$ .                  |
| c) $P(X = 1)$ .              | h) $P(XY \geq 2)$ .               |
| d) $P(Y = 2)$ .              | i) $P(Y \geq 2X)$ .               |
| e) $P(X = 0   Y = 2)$ .      | j) $P(X + Y \text{ sea impar})$ . |

358. Sean  $X$  y  $Y$  dos variables aleatorias continuas con función de densidad conjunta dada por la siguiente expresión:

$$f(x, y) = \begin{cases} 6x^2y & \text{si } 0 \leq x, y \leq 1, \\ 0 & \text{en otro caso.} \end{cases}$$

La gráfica de esta función se muestra en la Figura 4.7. Encuentre:

- |                                   |                            |
|-----------------------------------|----------------------------|
| a) $P(X \leq 1/2, Y \geq 1/2)$ .  | f) $P( X - Y  \leq 1/2)$ . |
| b) $P(Y \geq 1/2)$ .              | g) $P(XY < 1)$ .           |
| c) $P(X \leq 1/2   Y \geq 1/2)$ . | h) $P(Y \geq X^2)$ .       |
| d) $P(X + Y > 1)$ .               | i) $P(X^2 + Y^2 \leq 1)$ . |
| e) $P(Y > X)$ .                   | j) $P(Y \leq 4X(1 - X))$ . |

359. Se lanza un dado equilibrado dos veces consecutivas. Sea  $X$  el resultado del primer lanzamiento y sea  $Y$  el resultado del segundo lanzamiento. Encuentre la función de probabilidad de vector  $(X, X + Y)$ .

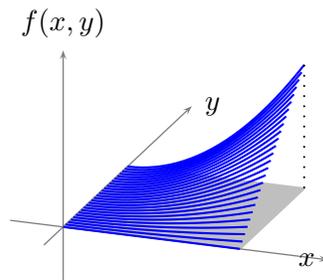


Figura 4.7

360. Demuestre que las siguientes funciones son de probabilidad:

$$a) f(x, y) = \begin{cases} 2^{-(x+y)} & \text{si } x, y = 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

$$b) f(x, y) = \begin{cases} 16 (1/3)^{x+2y} & \text{si } x, y = 1, 2, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

c) Sean  $n \in \mathbb{N}$  y  $p_1$  y  $p_2$  dos probabilidades. Para valores  $x, y = 0, 1, \dots, n$  tales que  $0 \leq x + y \leq n$  se define

$$f(x, y) = \frac{n!}{x! y! (n - x - y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}.$$

361. Demuestre que las siguientes funciones son de densidad:

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

$$b) f(x, y) = \begin{cases} 6 y^2 e^{-2x} & \text{si } 0 < y < 1, x > 0, \\ 0 & \text{en otro caso.} \end{cases}$$

$$c) f(x, y) = \begin{cases} 3xy(1-x) & \text{si } 0 < x < 1, 0 < y < 2, \\ 0 & \text{en otro caso.} \end{cases}$$

362. Encuentre el valor de la constante  $c$  para que cada una de las siguientes funciones sea de probabilidad.

$$\begin{aligned}
 a) \quad f(x, y) &= \begin{cases} \frac{c}{x! y!} & \text{si } x, y = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases} \\
 b) \quad f(x, y) &= \begin{cases} cxy & \text{si } 0 < x, y < 1, \\ 0 & \text{en otro caso.} \end{cases} \\
 c) \quad f(x, y, z) &= \begin{cases} c(x^2 + y^2 + z^2) & \text{si } -1 < x, y, z < 1, \\ 0 & \text{en otro caso.} \end{cases} \\
 d) \quad f(x, y, z) &= \begin{cases} c(x + y + z) & \text{si } 0 < x, y, z < 1, \\ 0 & \text{en otro caso.} \end{cases} \\
 e) \quad f(x_1, \dots, x_n) &= \begin{cases} c x_1 \cdots x_n & \text{si } 0 < x_i < 1, \quad i = 1, \dots, n, \\ 0 & \text{en otro caso.} \end{cases}
 \end{aligned}$$

363. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad

$$f(x, y) = \begin{cases} \frac{c^{x+y}}{x! y!} & \text{si } x, y = 0, 1, \dots \\ 0 & \text{en otro caso.} \end{cases}$$

- Encuentre el valor de la constante  $c$ .
- Calcule  $P(X + Y = n)$ ,  $n = 0, 1, \dots$

364. Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad

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

- Encuentre el valor de la constante  $c$ .
- Calcule  $P(X < Y)$ .
- Calcule  $P(X + Y > 1)$ .

365. Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad

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

- a) Encuentre el valor de la constante  $c$ .
- b) Calcule  $P(X^2 + Y^2 \leq r^2)$  para  $r \geq 0$ .
- c) Calcule  $P(X \leq \theta Y)$  para  $0 < \theta < 1$ .

366. Un dado equilibrado se lanza dos veces. Sea  $X$  la variable aleatoria que denota el menor de estos resultados. Encuentre la función de probabilidad de  $X$ .

### 4.3. Función de distribución conjunta

Además de la función de densidad o de probabilidad, existe la función de distribución para un vector  $(X, Y)$ , sea éste discreto o continuo. Su definición aparece a continuación y es muy semejante al caso unidimensional.

**Definición 4.4** La función de distribución del vector  $(X, Y)$ , denotada por  $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$ , se define de la siguiente manera:

$$F(x, y) = P(X \leq x, Y \leq y).$$

La pequeña coma que aparece en el lado derecho de esta igualdad significa la intersección de los eventos  $(X \leq x)$  y  $(Y \leq y)$ , es decir, el número  $F(x, y)$  es la probabilidad del evento  $(X \leq x) \cap (Y \leq y)$ . Más precisamente, esta función debe escribirse como  $F_{X,Y}(x, y)$ , pero recordemos que omitiremos los subíndices para mantener la notación simple. También aquí asociaremos la variable  $X$  con el valor  $x$ , y la variable  $Y$  con el valor  $y$ . A esta función se le conoce también con el nombre de **función de acumulación de probabilidad** del vector  $(X, Y)$ , y también se dice que es la **función de distribución conjunta** de las variables  $X$  y  $Y$ .

Enunciamos a continuación algunas propiedades que cumple toda función de distribución conjunta. Omitiremos la demostración de estas propiedades pues siguen el mismo tipo de ideas que en el caso unidimensional.

**Proposición 4.1** La función de distribución  $F(x, y)$  del vector aleatorio  $(X, Y)$  satisface las siguientes propiedades:

1.  $\lim_{x \rightarrow \infty} \lim_{y \rightarrow \infty} F(x, y) = 1$ .
2.  $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0$ .
3.  $F(x, y)$  es continua por la derecha en cada variable.
4.  $F(x, y)$  es una función monótona no decreciente en cada variable.
5. Para cualesquiera números  $a < b$ , y  $c < d$ , se cumple la desigualdad

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0.$$

Observe que las primeras cuatro propiedades son completamente análogas al caso unidimensional, y puede comprobarse geoméricamente que la quinta propiedad es idéntica a la probabilidad

$$P[(a < X \leq b) \cap (c < Y \leq d)].$$

que corresponde a la probabilidad de que el vector  $(X, Y)$  tome un valor dentro del rectángulo  $(a, b] \times (c, d]$  que se muestra en la Figura 4.8.

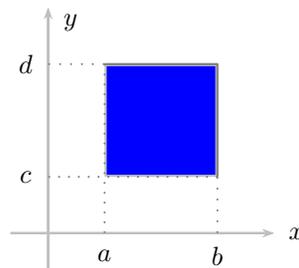


Figura 4.8

Recíprocamente, se dice que una función  $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$  es una **función de distribución conjunta** o **bivariada** si satisface las anteriores cinco

propiedades. En el caso continuo supondremos que la función de distribución bivariada  $F(x, y)$  puede expresarse de la siguiente forma:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du,$$

en donde  $f(x, y)$  es una función no negativa y corresponde a la función de densidad bivariada asociada. El concepto de función de distribución bivariada puede extenderse al caso de vectores multidimensionales de la siguiente forma.

**Definición 4.5** La función de distribución del vector aleatorio  $(X_1, \dots, X_n)$  es la función  $F(x_1, \dots, x_n) : \mathbb{R}^n \rightarrow [0, 1]$  dada por

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Regresemos ahora al caso bidimensional. Las funciones  $F(x, y)$  y  $f(x, y)$  son equivalentes y en nuestro caso es siempre posible encontrar una a partir de la otra. Explicaremos este procedimiento a continuación.

### De la función de densidad a la función de distribución

Conociendo la función de densidad  $f(x, y)$  se puede encontrar la función de distribución  $F(x, y)$  simplemente integrando en el caso continuo o sumando en el caso discreto. Para el caso continuo tenemos que

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du.$$

En el caso discreto se suman todos los valores de  $f(u, v)$  para valores de  $u$  menores o iguales a  $x$ , y valores de  $v$  menores o iguales a  $y$ , es decir,

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

**Ejemplo 4.8** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad  $f(x, y)$  dada por la siguiente tabla.

$x/y$	0	1
0	1/4	1/4
1	1/4	1/4

Para encontrar la función  $F(x, y)$  se necesita calcular la probabilidad  $P(X \leq x, Y \leq y)$  para cada par de números reales  $(x, y)$ . El plano Cartesiano  $\mathbb{R}^2$  puede ser dividido en cinco regiones y en cada una de ellas calcular la función de distribución, la cual adquiere la siguiente forma

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

Y cuyos valores se muestran en el plano Cartesiano que aparece en la Figura 4.9.

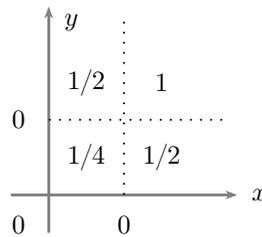


Figura 4.9

### De la función de distribución a la función de densidad

Recíprocamente, puede encontrarse la función de densidad  $f(x, y)$  a partir de la función de distribución  $F(x, y)$  de la siguiente forma: en el caso continuo

sabemos que  $f(x, y)$  y  $F(x, y)$  guardan la relación

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du,$$

y por el teorema fundamental del cálculo tenemos entonces que en los puntos  $(x, y)$  en donde  $f(x, y)$  es continua,

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

En el caso discreto,

$$f(x, y) = F(x, y) - F(x-, y-).$$

## Ejercicios

367. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad como indica la tabla de abajo. Encuentre y grafique la correspondiente función de distribución.

a) 

$x/y$	0	1
0	0	1/2
1	1/2	0

b) 

$x/y$	0	1	2	3
0	0	1/4	1/4	0
1	1/4	0	0	1/4

c) 

$x/y$	0	1	2
0	1/9	1/9	1/9
1	1/9	1/9	1/9
2	1/9	1/9	1/9

368. Sea  $(X, Y)$  un vector aleatorio continuo con función de probabilidad como se indica abajo. Encuentre y grafique la correspondiente función de distribución.

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

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

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

369. Sean  $F(x, y)$  y  $G(x, y)$  dos funciones de distribución bivariadas. Demuestre que para cualquier constante  $\lambda \in [0, 1]$  la siguiente función es de distribución:

$$(x, y) \mapsto \lambda F(x, y) + (1 - \lambda)G(x, y).$$

#### 4.4. Función de probabilidad marginal

Dada la función de densidad de un vector aleatorio, veremos ahora la forma de obtener la función de densidad de un subvector del vector aleatorio original. Veremos primero el caso bidimensional y después extenderemos las ideas al caso multidimensional discreto y continuo.

**Definición 4.6** Sea  $f(x, y)$  la función de densidad del vector aleatorio continuo  $(X, Y)$ . Se define la **función de densidad marginal** de la variable  $X$  como la siguiente integral

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Es decir, se integra simplemente respecto de la variable  $y$  para dejar como resultado una función que depende únicamente de  $x$ . Esta función resultante es la función de densidad marginal de  $X$ , y el subíndice 1 indica que se trata de la función de densidad marginal de la primera variable aleatoria del vector  $(X, Y)$ . De manera completamente análoga, la función de densidad marginal de la variable  $Y$  se obtiene integrando ahora respecto de la variable  $x$ , es

decir,

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Nuevamente, el subíndice 2 hace referencia a que esta función es la función de densidad de la segunda variable aleatoria del vector  $(X, Y)$ . En general, las funciones  $f_1(x)$  y  $f_2(y)$  son distintas, aunque hay ocasiones en que pueden ser iguales. Es inmediato verificar que estas funciones de densidad marginales son efectivamente funciones de densidad univariadas, pues son no negativas e integran uno.

**Ejemplo 4.9** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad dada por

$$f(x, y) = \begin{cases} 4xy & \text{si } 0 < x, y < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Es sencillo verificar que esta función es efectivamente una función de densidad bivariada pues es no negativa e integra uno:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^1 \int_0^1 4xy dx dy = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1.$$

Calcularemos ahora las funciones de densidad marginales  $f_1(x)$  y  $f_2(y)$ . Para  $x \in (0, 1)$ ,

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 4xy dy = 2x.$$

Por lo tanto,

$$f_1(x) = \begin{cases} 2x & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

De manera análoga, o por simetría,

$$f_2(y) = \begin{cases} 2y & \text{si } 0 < y < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Es inmediato comprobar que estas funciones son funciones de densidad univariadas. Observe que en este caso particular se cumple la identidad  $f(x, y) = f_1(x)f_2(y)$ , lo cual expresa el importante concepto de independencia de v.a.s que hemos mencionado antes. •

La definición de **función de probabilidad marginal** para vectores discretos involucra una suma en lugar de la integral, por ejemplo,

$$f_1(x) = \sum_y f(x, y),$$

de manera análoga se define la función de probabilidad marginal  $f_2(y)$ .

**Ejemplo 4.10** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad dada por

$$f(x, y) = \begin{cases} (x + 2y)/30 & \text{si } (x, y) \in \{1, 2, 3\} \times \{1, 2\}, \\ 0 & \text{otro caso.} \end{cases}$$

No es difícil comprobar que esta función es una función de probabilidad bivariada, es decir, es no negativa y suma uno,

$$\sum_{x=1}^3 \sum_{y=1}^2 \frac{x + 2y}{30} = 1.$$

Las funciones de probabilidad marginales  $f_1(x)$  y  $f_2(y)$  son

$$f_1(x) = \sum_{y=1}^2 f(x, y) = \begin{cases} 8/30 & \text{si } x = 1, \\ 10/30 & \text{si } x = 2, \\ 12/30 & \text{si } x = 3, \\ 0 & \text{en otro caso.} \end{cases}$$

y

$$f_2(y) = \sum_{x=1}^3 f(x, y) = \begin{cases} 12/30 & \text{si } y = 1, \\ 18/30 & \text{si } y = 2, \\ 0 & \text{en otro caso.} \end{cases}$$

Estas funciones son funciones de probabilidad univariadas. •

Un poco más generalmente, la función de densidad marginal de la variable  $X_1$  a partir de la función de densidad del vector  $(X_1, \dots, X_n)$  es, en el caso continuo,

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_2 \cdots dx_n.$$

De manera análoga se puede obtener la función de densidad marginal de cualquiera de las variables que componen el vector multidimensional. Y también de manera similar se pueden calcular estas densidades marginales de vectores que son subconjuntos del vector original, por ejemplo, la función de densidad marginal del vector  $(X_1, X_2)$  a partir de  $(X_1, \dots, X_n)$  es

$$f_{1,2}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_3 \cdots dx_n.$$

Otro aspecto interesante sobre estas funciones es que pueden existir distintas funciones de densidad conjuntas que producen las mismas funciones de densidad marginales. En el Ejercicio 371 se muestra esta situación.

## Ejercicios

370. Encuentre las funciones de probabilidad marginales  $f_1(x)$  y  $f_2(y)$  para cada una de las siguientes funciones de probabilidad conjuntas.

a) 

$x/y$	0	1
0	1/16	5/16
1	4/16	6/16

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

c) 
$$f(x, y) = \frac{n!}{x! y! (n - x - y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y},$$

en donde  $n \in \mathbb{N}$ ,  $p_1$  y  $p_2$  son dos probabilidades y  $x, y = 0, 1, \dots, n$  tales que  $0 \leq x + y \leq n$ .

371. **Varias conjuntas, mismas marginales.** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad dada por la tabla que aparece abajo. Compruebe que para cualquier valor de los parámetros  $\theta$  y  $p$  tales que  $0 \leq p \leq 1/2$  y  $(1 - 2p)/(1 - p) \leq \theta \leq 1$ , las correspondientes funciones de probabilidad marginales son  $\text{Ber}(p)$ .

$x/y$	0	1
0	$\theta(1-p)$	$(1-\theta)(1-p)$
1	$(1-\theta)(1-p)$	$p - (1-\theta)(1-p)$

## 4.5. Función de distribución marginal

Ahora veremos la forma de obtener la función de distribución de un subvector a partir de la función de distribución del vector aleatorio original. Nuevamente consideraremos primero el caso bidimensional y después veremos el caso de dimensiones mayores.

**Definición 4.7** Sea  $(X, Y)$  un vector aleatorio, continuo o discreto, con función de distribución  $F(x, y)$ . La función de distribución marginal de la variable  $X$  se define como la función de una variable

$$F_1(x) = \lim_{y \rightarrow \infty} F(x, y),$$

La correspondiente función de distribución marginal de la variable  $Y$  se define como la función

$$F_2(y) = \lim_{x \rightarrow \infty} F(x, y).$$

Estos límites siempre existen pues la función de distribución conjunta es acotada y no decreciente en cada variable. No es difícil comprobar que estas funciones de distribución marginales son efectivamente funciones de distribución univariadas. Para un vector de dimensión tres  $(X, Y, Z)$ , a partir de  $F_{X,Y,Z}(x, y, z)$  y tomando los límites necesarios, pueden obtenerse, por ejemplo, las funciones de distribución marginales  $F_{X,Y}(x, y)$ ,  $F_{X,Z}(x, z)$ ,  $F_X(x)$ . En efecto, para el último caso tenemos que

$$F_X(x) = \lim_{y \rightarrow \infty} \lim_{z \rightarrow \infty} F_{X,Y,Z}(x, y, z).$$

Más generalmente, a partir de la función de distribución de un vector  $(X_1, \dots, X_n)$  se puede obtener de manera análoga la función de distribución de cualquier subvector.

Ahora que conocemos la forma de obtener las funciones de densidad y de distribución marginales a partir de las correspondientes funciones conjuntas, podemos enunciar con precisión el concepto de independencia entre variables aleatorias. Veremos en la siguiente sección este importante concepto el cual resulta ser una hipótesis recurrente en los procedimientos de la estadística matemática y otras áreas de aplicación de la probabilidad.

## Ejercicios

372. Sean  $X$  y  $Y$  dos variables aleatorias discretas con función de probabilidad conjunta dada por la tabla que aparece abajo. Encuentre la función de distribución conjunta  $F_{X,Y}(x, y)$  y a partir de ella encuentre las funciones de distribución marginales  $F_X(x)$  y  $F_Y(y)$ .

$x/y$	0	1
0	1/8	2/8
1	2/8	3/8

373. Encuentre las funciones de distribución marginales  $F_1(x)$  y  $F_2(y)$  para cada una de las siguientes funciones de distribución conjuntas. En cada caso grafique  $F(x, y)$ ,  $F_1(x)$  y  $F_2(y)$ .

- a) Si  $a < b$  y  $c < d$ , entonces

$$F(x, y) = \begin{cases} 0 & \text{si } x < a \text{ ó } y < c, \\ 1/2 & \text{si } a \leq x < b, c \leq y < d, \\ 3/4 & \text{si } a \leq x < b, y \geq d, \\ 3/4 & \text{si } x \geq b, c \leq y < d, \\ 1 & \text{si } x \geq b, y \geq d. \end{cases}$$

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

## 4.6. Independencia de variables aleatorias

Sea  $X_1, \dots, X_n$  una colección de variables aleatorias con función de distribución conjunta  $F(x_1, \dots, x_n)$ . Suponga que las respectivas funciones de distribución marginales son  $F_1(x_1), \dots, F_n(x_n)$ .

**Definición 4.8** Se dice que las variables aleatorias  $X_1, \dots, X_n$  son **independientes** si para cualesquiera números reales  $x_1, \dots, x_n$  se cumple la igualdad

$$F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n).$$

Alternativamente, puede definirse la independencia en términos de la función de densidad como sigue

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Nuevamente esta igualdad debe verificarse para cualesquiera valores  $x_1, \dots, x_n$ . Puede demostrarse que las dos condiciones anteriores son equivalentes. En el caso particular cuando las variables aleatorias son discretas, la condición de independencia se escribe de la forma siguiente: para cualesquiera números  $x_1, \dots, x_n$ ,

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

**Ejemplo 4.11** Sean  $X$  y  $Y$  dos variables aleatorias continuas con función de densidad conjunta  $f(x, y)$  dada por

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

Pueden calcularse las funciones de densidad marginales y encontrar que

$$f_1(x) = \begin{cases} e^{-x} & \text{si } x > 0, \\ 0 & \text{otro caso,} \end{cases} \quad y \quad f_2(y) = \begin{cases} e^{-y} & \text{si } y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

Se verifica entonces que  $f(x, y) = f_1(x)f_2(y)$  para cualesquiera números reales  $x$  y  $y$ , y ello demuestra la independencia de las variables  $X$  y  $Y$ . •

**Ejemplo 4.12** Sean  $X$  y  $Y$  dos variables aleatorias discretas con función de probabilidad  $f(x, y)$  dada por

$$f(x, y) = \begin{cases} 1/4 & \text{si } x, y \in \{0, 1\}, \\ 0 & \text{otro caso.} \end{cases}$$

Las funciones de probabilidad marginales son

$$f_1(x) = \begin{cases} 1/2 & \text{si } x \in \{0, 1\}, \\ 0 & \text{otro caso,} \end{cases} \quad y \quad f_2(y) = \begin{cases} 1/2 & \text{si } y \in \{0, 1\}, \\ 0 & \text{otro caso.} \end{cases}$$

Por lo tanto  $f(x, y) = f_1(x)f_2(y)$  para cualesquiera números reales  $x$  y  $y$ , y se concluye entonces que  $X$  y  $Y$  son independientes. •

Adicionalmente tenemos la siguiente extensión del concepto de independencia de variables aleatorias.

**Definición 4.9** Se dice que un conjunto infinito de variables aleatorias es **independiente** si cualquier subconjunto finito de él lo es.

Este es el sentido en el que una sucesión infinita de variables aleatorias independientes debe entenderse. Tal hipótesis aparece, por ejemplo, en el enunciado de la ley de los grandes números y el teorema central del límite que estudiaremos más adelante.

## Ejercicios

374. Sean  $X$  y  $Y$  dos variables aleatorias discretas cuya función de probabilidad conjunta admite la descomposición

$$P(X = x, Y = y) = g(x)h(y),$$

para ciertas funciones  $g(x)$  y  $h(y)$ .

- a) Expresar  $P(X = x)$  y  $P(Y = y)$  en términos de  $g(x)$  y  $h(y)$ .

- b) Demuestre que  $(\sum_x g(x)) (\sum_y h(y)) = 1$ .
- c) Demuestre que  $X$  y  $Y$  son independientes.
375. Sean  $X$  y  $Y$  dos variables aleatorias independientes con distribución geométrica de parámetros  $p$  y  $q$  respectivamente. Calcule:
- a)  $P(X = Y)$ .
- b)  $P(X \leq Y)$ .
376. Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias independientes cada una con la misma distribución  $\text{Ber}(p)$ , e independientes de otra variable aleatoria  $N$  con distribución  $\text{Poisson}(\lambda)$ . Demuestre que

$$X = \sum_{i=1}^N X_i \sim \text{Poisson}(\lambda p).$$

Cuando  $N = 0$  la suma es vacía y se define ésta como cero. Si  $N$  representa el número de delitos ocurridos de los cuales sólo la fracción  $p$  son reportados a la autoridad, entonces  $X$  representa el número de delitos reportados.

377. El siguiente ejemplo muestra que la condición  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$  no es suficiente para concluir que  $X$  y  $Y$  son independientes. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad dada por la siguiente tabla

$x/y$	-1	0	1
-1	1/8	0	1/8
0	0	1/2	0
1	1/8	0	1/8

Compruebe que se cumple la igualdad  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$  y que  $X$  y  $Y$  no son independientes.

378. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad  $f(x, y)$  dada por la siguiente tabla.

$x \setminus y$	0	1	2
0	0.1	0.2	0.1
1	0.1	0	0.1
2	0.1	0.1	0.2

- a) Grafique  $f(x, y)$  y demuestre que efectivamente se trata de una función de probabilidad.
- b) Calcule y grafique las densidades marginales  $f_1(x)$  y  $f_2(y)$ . Verifique que ambas son efectivamente funciones de probabilidad.
- c) Calcule y grafique la función de distribución conjunta  $F(x, y)$ .
- d) Calcule y grafique las distribuciones marginales  $F_1(x)$  y  $F_2(y)$ . Verifique que ambas son efectivamente funciones de distribución.
- e) ¿Son  $X$  y  $Y$  independientes?
379. Sea  $(X, Y)$  un vector aleatorio continuo con distribución uniforme en el cuadrado  $(-1, 1) \times (-1, 1)$ .
- a) Grafique  $f(x, y)$  y demuestre que efectivamente se trata de una función de densidad.
- b) Calcule y grafique las densidades marginales  $f_1(x)$  y  $f_2(y)$ . Verifique que ambas son efectivamente funciones de densidad.
- c) Calcule y grafique la función de distribución conjunta  $F(x, y)$ .
- d) Calcule y grafique las distribuciones marginales  $F_1(x)$  y  $F_2(y)$ . Verifique que ambas son efectivamente funciones de distribución.
- e) ¿Son  $X$  y  $Y$  independientes?
380. Determine si las siguientes funciones de densidad corresponden a variables aleatorias independientes.

a) 

$x/y$	0	1
0	1/4	1/4
1	1/4	1/4

b) 

$x/y$	0	1	2
0	1/30	2/30	3/30
1	4/30	5/60	6/30
2	3/30	3/30	3/30

c) 

$x/y$	0	1	2
0	1/60	2/60	3/60
1	2/60	4/60	6/60
2	3/60	6/60	9/60
3	4/60	8/60	12/60

$$\begin{aligned}
 d) \quad f(x, y) &= \begin{cases} 4xy & \text{si } 0 < x, y < 1, \\ 0 & \text{otro caso.} \end{cases} \\
 e) \quad f(x, y) &= \begin{cases} x + y & \text{si } 0 < x, y < 1, \\ 0 & \text{otro caso.} \end{cases} \\
 f) \quad f(x, y, z) &= \begin{cases} 8xyz & \text{si } 0 < x, y, z < 1, \\ 0 & \text{otro caso.} \end{cases} \\
 g) \quad f(x, y, z) &= \begin{cases} x^2 + y^2 + z^2 & \text{si } 0 < x, y, z < 1, \\ 0 & \text{otro caso.} \end{cases} \\
 h) \quad f(x_1, \dots, x_n) &= \begin{cases} 2^n x_1 \cdots x_n & \text{si } 0 < x_i < 1, i = 1, \dots, n, \\ 0 & \text{otro caso.} \end{cases}
 \end{aligned}$$

## 4.7. Distribución condicional

Recordemos que la probabilidad condicional de un evento  $A$  dado un evento  $B$  está dada por

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

Esta definición puede extenderse al caso de funciones de probabilidad o de densidad y también para el caso de funciones de distribución. En esta sección enunciaremos tales extensiones.

**Definición 4.10** Sea  $(X, Y)$  un vector aleatorio discreto (o continuo) y con función de probabilidad (o de densidad)  $f_{X,Y}(x, y)$ . Sea  $y$  un valor de la variable  $Y$  tal que  $f_Y(y) \neq 0$ . A la función  $x \mapsto f_{X|Y}(x|y)$  definida a continuación se le llama la **función de probabilidad** (o **densidad**) de  $X$  dado que  $Y = y$ ,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \quad (4.2)$$

Observe que a la función dada por (4.2) se le considera como una función de  $x$  y que el valor de  $y$  es fijo y puede considerársele como un parámetro de dicha función, es decir, para cada valor fijo de  $y$  se tiene una función diferente. En el caso discreto la expresión (4.2) es efectivamente la definición de probabilidad condicional

$$f_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)},$$

sin embargo, recordemos que en el caso continuo las expresiones  $f_{X,Y}(x, y)$  y  $f_Y(y)$  no son probabilidades. Sumando o integrando sobre los posibles valores  $x$ , es inmediato comprobar que la función dada por (4.2) es efectivamente una función de probabilidad o de densidad. Observe además que cuando  $X$  y  $Y$  son independientes,

$$f_{X|Y}(x|y) = f_X(x).$$

**Ejemplo 4.13** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad dada por la siguiente tabla:

$x/y$	0	1	2	3
0	0.1	0.1	0.2	0.1
1	0.1	0.2	0.1	0.1

Calcularemos la función de probabilidad condicional  $f_{X|Y}(x|y)$  para  $y = 1$ . Sumando las probabilidades de la columna correspondiente a  $y = 1$  se encuentra que  $f_Y(1) = 0.3$ . Por lo tanto, siguiendo la fórmula de la expresión (4.2) tenemos que

$$f_{X|Y}(x|1) = \frac{f_{X,Y}(x, 1)}{f_Y(1)} = \begin{cases} 1/3 & \text{si } x = 0, \\ 2/3 & \text{si } x = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

De manera análoga pueden calcularse  $f_{X|Y}(x|0)$ ,  $f_{X|Y}(x|2)$ ,  $f_{X|Y}(x|3)$ , y también  $f_{Y|X}(y|0)$  y  $f_{Y|X}(y|1)$ . ¿Puede usted encontrar estas funciones de probabilidad? •

**Ejemplo 4.14** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad dada por

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

Calcularemos la función de probabilidad condicional  $f_{X|Y}(x|y)$  para cada  $y$  en el intervalo  $(0, 1)$ . Integrando sobre  $x$  tenemos que la función de densidad marginal de  $Y$  es

$$f_Y(y) = \begin{cases} \frac{1}{2} + y & \text{si } 0 < y < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo tanto, para cada  $y \in (0, 1)$  fijo, la función de densidad condicional de  $X$  dado  $Y = y$  está dada por

$$x \mapsto f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \begin{cases} \frac{x+y}{1/2+y} & \text{si } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

De manera análoga puede calcularse  $y \mapsto f_{Y|X}(y|x)$  para cada  $x \in (0, 1)$ . •

La fórmula (4.2) puede extenderse de manera análoga al caso de vectores de dimensión mayor. Por ejemplo, para un vector aleatorio de dimensión tres  $(X, Y, Z)$  pueden calcularse funciones de densidad condicionales como  $f_{X|Y,Z}(x|y, z)$  o  $f_{X,Z|Y}(x, z|y)$ .

Veamos ahora la extensión al caso de funciones de distribución.

**Definición 4.11** Sea  $(X, Y)$  un vector aleatorio con función de probabilidad o densidad  $f_{X,Y}(x, y)$ . Sea  $y$  un valor de  $Y$  tal que  $f_Y(y) \neq 0$ . La **función de distribución condicional** de  $X$  dado  $Y = y$  es la función

$$F_{X|Y}(x|y) = \begin{cases} \sum_{u \leq x} f_{X|Y}(u|y) & \text{en el caso discreto,} \\ \int_{-\infty}^x f_{X|Y}(u|y) du & \text{en el caso continuo.} \end{cases}$$



- a) Encuentre la función de probabilidad conjunta de  $X$  y  $Y$ .  
 b) Calcule las funciones  $f_{Y|X}(y|x=3)$  y  $f_{X|Y}(x|y=3)$ .

## 4.8. Esperanza condicional

Definiremos a continuación el valor esperado de una variable aleatoria dado que un evento ha ocurrido para otra variable aleatoria cuando se conoce la distribución conjunta de las dos variables.

**Definición 4.12** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad  $f_{X,Y}(x, y)$  y sea  $y$  un valor tal que  $f_Y(y) \neq 0$ . La **esperanza condicional** de  $X$  dado  $Y = y$  es la esperanza de la función de densidad condicional  $f_{X|Y}(x|y)$ , cuando existe, es decir,

$$E(X | Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx.$$

Efectuando un cambio en el orden de las integrales es inmediato comprobar que

$$E(X) = \int_{-\infty}^{\infty} E(X | Y = y) f_Y(y) dy,$$

cuya forma recuerda el teorema de probabilidad total, pero esta vez en términos de esperanzas. En el caso cuando el vector  $(X, Y)$  es discreto la definición es análoga:

$$\begin{aligned} E(X | Y = y) &= \sum_x x f_{X|Y}(x|y) \\ &= \sum_x x P(X = x | Y = y), \end{aligned}$$

suponiendo nuevamente que  $f_Y(y) \neq 0$  y que la suma indicada es absolutamente convergente. Y nuevamente, efectuando un cambio en el orden de las

sumas se encuentra la expresión

$$E(X) = \sum_y E(X | Y = y) P(Y = y).$$

En cualquier caso observe además que cuando  $X$  y  $Y$  son independientes,

$$E(X | Y = y) = E(X).$$

## Ejercicios

384. Un experimento consiste en lanzar un dado equilibrado repetidas veces hasta obtener alguno de los resultados por segunda vez. Encuentre el número esperado de lanzamientos en este experimento.

385. Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad como aparece en la tabla de abajo. Calcule  $E(X | Y = 0)$ ,  $E(X | Y = 1)$  y  $E(X | Y = 2)$ . Verifique además que se cumple la identidad

$$E(X) = \sum_{y=0}^2 E(X | Y = y) P(Y = y).$$

$x/y$	0	1	2
0	1/10	1/4	1/10
1	1/10	1/4	2/10

386. Se lanza un dado equilibrado dos veces consecutivas. Sea  $X$  el resultado del primer lanzamiento y sea  $Y$  el resultado del segundo lanzamiento. Calcule

$$E(X | X + Y = 6).$$

387. Para cada  $y$  en  $(0, 1)$  calcule la esperanza condicional  $E(X | Y = y)$  cuando  $X$  y  $Y$  tienen función de densidad conjunta

$$f_{X,Y}(x, y) = \begin{cases} 12x^2 & \text{si } 0 < x < y < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

388. Sea  $X$  con distribución  $\text{Poisson}(\Lambda)$  en donde  $\Lambda$  es una variable aleatoria con distribución  $\text{unif}\{1, 2\}$ . Condicionando sobre el valor de  $\Lambda$  encuentre  $E(X)$ .
389. Sea  $X$  con distribución  $\text{bin}(N, p)$  en donde  $N$  es una variable aleatoria con distribución  $\text{unif}\{1, 2\}$ . Condicionando sobre el valor de  $N$  encuentre  $E(X)$ .
390. Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias independientes e idénticamente distribuídas con media y varianza finitas, e independientes de otra variable  $N$  con valores  $1, 2, \dots$  y con esperanza y varianza finitas. Defina la variable aleatoria

$$S = \sum_{i=1}^N X_i.$$

Condicionando sobre el valor de  $N$  y usando las hipótesis de independencia demuestre que:

- a)  $E(S) = E(N) E(X)$ .
- b)  $\text{Var}(S) = \text{Var}(N) E^2(X) + \text{Var}(X) E(N)$ .

## 4.9. Covarianza

Sean  $X$  y  $Y$  dos variables aleatorias con esperanza finita y con función de densidad o de probabilidad conjunta  $f(x, y)$ . La covarianza entre  $X$  y  $Y$  es un número real que se denota por  $\text{Cov}(X, Y)$  y se define como la esperanza de la variable aleatoria  $(X - E(X))(Y - E(Y))$ .

**Definición 4.13** La covarianza entre las variables aleatorias  $X$  y  $Y$  es el número real

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Como veremos más adelante, la covarianza está estrechamente relacionada con otro concepto que se define para dos variables aleatorias llamado coeficiente de correlación, y para el cual se cuenta con una interpretación bastante clara. Dejaremos entonces la interpretación de la covarianza en términos del coeficiente de correlación. Explicaremos ahora la forma de calcular la covarianza según la definición anterior. Cuando  $X$  y  $Y$  son variables aleatorias discretas la covarianza se calcula de la forma siguiente:

$$\text{Cov}(X, Y) = \sum_{x,y} (x - E(X))(y - E(Y)) f(x, y),$$

en donde, observe, la suma es doble, se suma sobre todos los posibles valores  $x$  y también sobre todos los posibles valores  $y$ . En el caso cuando las variables aleatorias son continuas se tiene que

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f(x, y) dx dy.$$

Desarrollando el producto que aparece en la definición de covarianza y aplicando la linealidad de la esperanza se encuentra que la covarianza puede calcularse también como indica la siguiente fórmula:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Por otro lado, a partir de la definición misma de covarianza, o a partir de la fórmula recién enunciada, es inmediato observar que la covarianza es simétrica, es decir,  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ . Otra propiedad interesante y fácil de obtener se encuentra cuando se calcula la covarianza entre una variable aleatoria  $X$  y ella misma, en este caso la covarianza se reduce a la varianza de  $X$ . Es posible demostrar también que si  $X$  y  $Y$  son independientes, entonces  $\text{Cov}(X, Y) = 0$ . El recíproco es en general falso, es decir, el hecho de que la covarianza sea cero no implica necesariamente que las variables aleatorias en cuestión sean independientes. Por último, recordemos que hemos mencionado que la varianza de la suma de dos variables aleatorias no es, en general, la suma de las varianzas, sin embargo se cuenta con la siguiente fórmula general la cual puede ser encontrada a partir de la definición de varianza y se deja como ejercicio al lector.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y).$$

En la siguiente sección demostraremos además que

$$-\sqrt{\text{Var}(X)\text{Var}(Y)} \leq \text{Cov}(X, Y) \leq +\sqrt{\text{Var}(X)\text{Var}(Y)}.$$

## Ejercicios

391. Calcule la covarianza entre  $X$  y  $Y$  cuando estas variables tienen distribución conjunta como se indica en cada inciso:

a)  $f(x, y)$  está dada por la siguiente tabla:

$x/y$	0	1
0	1/4	1/4
1	1/4	1/4

$$b) f(x, y) = \begin{cases} 3x & \text{si } 0 < x < y < 1, \\ 3y & \text{si } 0 < y < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

392. **Propiedades de la covarianza.** Demuestre con detalle las siguientes propiedades de la covarianza:

a)  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ .

b)  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ , (simetría).

c)  $\text{Cov}(X, c) = \text{Cov}(c, X) = 0$ ,  $c$  constante.

d)  $\text{Cov}(cX, Y) = \text{Cov}(X, cY) = c \text{Cov}(X, Y)$ ,  $c$  constante.

e)  $\text{Cov}(X + c, Y) = \text{Cov}(X, Y + c) = \text{Cov}(X, Y)$ ,  $c$  constante.

f)  $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$ ,

Esta propiedad junto con la anterior y la simetría establecen que la covarianza es una función lineal en cada variable.

g)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$ .

h) Si  $X$  y  $Y$  son independientes entonces  $\text{Cov}(X, Y) = 0$ .

En consecuencia, cuando se cumple la hipótesis de independencia se tiene que  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ . Es útil observar también que la propiedad enunciada proporciona un mecanismo

para comprobar que dos variables aleatorias no son independientes pues, si sabemos que  $\text{Cov}(X, Y) \neq 0$  podemos entonces concluir que  $X$  y  $Y$  no son independientes.

*i)* En general,  $\text{Cov}(X, Y) = 0 \not\Rightarrow X, Y$  independientes.

393. Sean  $X$  y  $Y$  variables aleatorias con valores en el intervalo  $[a, b]$ . Demuestre que

$$|\text{Cov}(X, Y)| \leq \frac{1}{4}(b - a)^2.$$

## 4.10. Coeficiente de correlación

Habiendo definido la covarianza podemos ahora dar la definición del coeficiente de correlación entre dos variables aleatorias. Supondremos que tales variables aleatorias tienen esperanza y varianza finitas.

**Definición 4.14** El coeficiente de correlación entre las variables aleatorias  $X$  y  $Y$  con varianzas finitas distintas de cero se define como el número

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

Al número  $\rho(X, Y)$  se le denota también por  $\rho_{X, Y}$ , en donde  $\rho$  es la letra griega ro. El lector puede observar inmediatamente que la diferencia entre la covarianza y el coeficiente de correlación radica únicamente en que este último se obtiene al dividir la covarianza por el producto de las desviaciones estándar de las variables aleatorias. Puede demostrarse que este cambio de escala tiene como consecuencia que el coeficiente de correlación tome como valor máximo 1, y como valor mínimo  $-1$ , es decir,

**Proposición 4.2** El coeficiente de correlación entre dos variables aleatorias  $X$  y  $Y$  con varianzas finitas distintas de cero satisface las siguientes dos desigualdades:

$$-1 \leq \rho(X, Y) \leq 1.$$

Véase el ejercicio 395 para una demostración de este resultado. Explicaremos ahora la interpretación del coeficiente de correlación. Cuando  $X$  y  $Y$  son tales que  $\rho(X, Y) = 1$ , entonces existen constantes  $a$  y  $b$ , con  $a$  positiva tales que  $Y = aX + b$ , es decir, se puede establecer una dependencia lineal directa entre las dos variables aleatorias. En el otro caso extremo, cuando  $\rho(X, Y) = -1$ , entonces nuevamente existen constantes  $a$  y  $b$ , pero ahora con  $a$  negativa, tales que  $Y = aX + b$ . De nuevo, se trata de una relación lineal entre las dos variables aleatorias pero ahora tal relación es inversa en el sentido de que cuando una de las variables aleatorias crece la otra decrece. De esta forma el coeficiente de correlación es una medida del grado de dependencia lineal entre dos variables aleatorias. Existen varias formas en que dos variables aleatorias pueden depender una de otra, el coeficiente de correlación no mide todas estas dependencias, únicamente mide la dependencia de tipo lineal. Así, hemos mencionado que cuando el coeficiente de correlación es  $+1$ , ó  $-1$ , la dependencia lineal es exacta. Como en el caso de la covarianza, puede demostrarse que si dos variables aleatorias son independientes, entonces el coeficiente de correlación es cero, y nuevamente, el recíproco es en general falso, es decir, la condición de que el coeficiente de correlación sea cero no es suficiente para garantizar que las variables aleatorias sean independientes, excepto en el caso cuando las variables tienen distribución conjunta normal.

## Ejercicios

394. Calcule el coeficiente de correlación entre  $X$  y  $Y$  cuando estas variables tienen distribución conjunta como se indica en cada inciso:

a)  $f(x, y)$  está dada por la siguiente tabla:

$x/y$	0	1
0	1/4	1/4
1	1/4	1/4

$$b) f(x, y) = \begin{cases} 3x & \text{si } 0 < x < y < 1, \\ 3y & \text{si } 0 < y < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

395. Demuestre las siguientes afirmaciones y a partir de ellas demuestre que el coeficiente de correlación  $\rho(X, Y)$  únicamente toma valores en el intervalo  $[-1, 1]$ .

$$0 \leq \text{Var}\left(\frac{X}{\sqrt{\text{Var}(X)}} + \frac{Y}{\sqrt{\text{Var}(Y)}}\right) = 2(1 + \rho(X, Y)).$$

$$0 \leq \text{Var}\left(\frac{X}{\sqrt{\text{Var}(X)}} - \frac{Y}{\sqrt{\text{Var}(Y)}}\right) = 2(1 - \rho(X, Y)).$$

396. **Otras propiedades del coeficiente de correlación.** Demuestre las siguientes propiedades.

a)  $\rho(X, X) = 1$ .

b)  $\rho(X, -X) = -1$ .

c)  $\rho(X, Y) = \rho(Y, X)$ .

d)  $\rho(cX, Y) = \rho(X, cY) = \text{signo}(c)\rho(X, Y)$ ,  $c \neq 0$  constante.

e)  $\rho(X + c, Y) = \rho(X, Y + c) = \rho(X, Y)$ ,  $c$  constante.

f)  $\rho(X, aX + b) = \text{signo}(a)$ ,  $a \neq 0$ ,  $b$  constantes.



## Capítulo 5

# Teoremas límite

En este último capítulo estudiaremos dos de los teoremas límite más importantes en la probabilidad: la ley de los grandes números y el teorema central del límite. Mostraremos algunos ejemplos del uso y aplicación de estos resultados.

### 5.1. Desigualdad de Chebyshev

Este resultado es de carácter teórico y proporciona una cota superior para la probabilidad de que una variable aleatoria tome un valor que diste de su media en más de un factor de su desviación estándar.

**Proposición 5.1 (Desigualdad de Chebyshev)** Sea  $X$  una variable aleatoria con media  $\mu$  y varianza finita  $\sigma^2$ . Para cualquier número real  $\epsilon > 0$ ,

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}. \quad (5.1)$$

*Demostración.* Supongamos primero que  $X$  es continua con función de

densidad  $f(x)$ . Entonces

$$\begin{aligned}
 \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\
 &\geq \int_{|x - \mu| \geq \epsilon} (x - \mu)^2 f(x) dx \\
 &\geq \epsilon^2 \int_{|x - \mu| \geq \epsilon} f(x) dx \\
 &= \epsilon^2 P(|X - \mu| \geq \epsilon).
 \end{aligned}$$

Despejando la probabilidad encontrada se obtiene el resultado. La demostración es enteramente análoga en el caso cuando  $X$  es discreta, para ello debe reemplazarse la integral por el símbolo de suma. •

Observemos que el parámetro  $\epsilon$  que aparece en la desigualdad de Chebyshev debe ser en realidad estrictamente mayor a  $\sigma^2$  pues de lo contrario, si  $0 < \epsilon \leq \sigma^2$ , entonces  $\sigma^2/\epsilon^2 \geq 1$  y tal cantidad no proporciona ninguna información útil como cota superior para una probabilidad. Es también interesante observar que la desigualdad de Chebyshev es óptima en el sentido de que sin hipótesis adicionales puede alcanzarse la cota superior. En el Ejercicio 401 se pide comprobar este hecho en un caso particular. Haremos uso de la desigualdad de Chebyshev más adelante para demostrar la ley débil de los grandes números.

## Ejercicios

397. Bajo las mismas condiciones y notación del enunciado de la desigualdad de Chebyshev, demuestre directamente las siguientes versiones:

$$a) P(|X - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}.$$

$$b) P(|X - \mu| \geq \epsilon\sigma) \leq \frac{1}{\epsilon^2}.$$

$$c) P(|X - \mu| \leq \epsilon\sigma) \geq 1 - \frac{1}{\epsilon^2}.$$

398. **Desigualdad de Markov.** Demuestre cada uno de los siguientes resultados. A cualquier de ellos se le llama desigualdad de Markov.

- a) Sea  $X$  una variable aleatoria no negativa y con esperanza finita  $\mu$ . Demuestre que para cualquier constante  $\epsilon > 0$ ,

$$P(X \geq \epsilon) \leq \frac{\mu}{\epsilon}. \quad (5.2)$$

Sugerencia: utilice la misma técnica que se usó para demostrar la desigualdad de Chebyshev, esta vez empiece escribiendo la definición de esperanza para una variable aleatoria continua no negativa.

- b) Sea  $X$  una variable aleatoria no negativa y con  $n$ -ésimo momento finito. Demuestre que para cualquier constante  $\epsilon > 0$ ,

$$P(X \geq \epsilon) \leq \frac{E(X^n)}{\epsilon^n}.$$

- c) Sea  $X$  una variable aleatoria y sea  $\varphi : \mathbb{R} \rightarrow [0, \infty)$  una función monótona no decreciente tal que  $E(\varphi(X)) < \infty$ . Demuestre que para cualquier constante  $\epsilon > 0$ ,

$$P(X \geq \epsilon) \leq \frac{E(\varphi(X))}{\varphi(\epsilon)}.$$

399. Sea  $X$  una variable aleatoria con f.g.m.  $M_X(t)$ . Demuestre que para  $t > 0$  y  $x > 0$ ,

$$P(X \geq x) \leq e^{-tx} M_X(t).$$

400. [Markov  \$\Rightarrow\$  Chebyshev](#). Use la desigualdad de Markov (5.2) aplicada a la variable aleatoria no negativa  $|X - \mu|^2$  para demostrar la desigualdad de Chebyshev (5.1).

401. [La desigualdad de Chebyshev es óptima](#). Sea  $X$  una variable aleatoria discreta con función de probabilidad

$$f(x) = \begin{cases} 1/18 & \text{si } x = -1, 1, \\ 16/18 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases}$$

- a) Calcule la esperanza  $\mu$  y la varianza  $\sigma^2$  de esta v.a.

- b) Ahora calcule exactamente  $P(|X - \mu| \geq 3\sigma)$  y compruebe que esta cantidad coincide con la cota superior dada por la desigualdad de Chebyshev.

Este resultado demuestra que sin hipótesis adicionales la cota superior dada por la desigualdad de Chebyshev es óptima, es decir, no puede establecerse una cota superior más pequeña.

402. Sea  $X$  una variable aleatoria con distribución  $N(\mu, \sigma^2)$ .
- a) Use la desigualdad de Chebyshev para estimar el valor mínimo de  $k$  de tal modo que la probabilidad de que  $X$  tome un valor entre  $\mu - k\sigma$  y  $\mu + k\sigma$  sea al menos 0.95.
- b) Use la tabla de la distribución normal para encontrar el valor exacto de  $k$  que cumpla la condición del inciso anterior.
403. Sea  $\Phi(x)$  la función de distribución  $N(0, 1)$ . Use la desigualdad de Chebyshev para demostrar que para cualquier  $x > 0$ ,

$$2x^2\Phi(-x) \leq 1.$$

404. Use la distribución  $\exp(\lambda)$  y la desigualdad de Chebyshev para demostrar que para cualquier número real  $x \geq 1$ ,

$$e^{-(x+1)} \leq \frac{1}{x^2}.$$

## 5.2. Convergencia de variables aleatorias

Sabemos que una sucesión numérica  $x_1, x_2, \dots$  es convergente a un número  $x$  si para cualquier  $\epsilon > 0$  existe un número natural  $N$  a partir del cual los elementos de la sucesión se encuentran cercanos al número  $x$ , es decir, para  $n \geq N$ ,

$$|x_n - x| < \epsilon.$$

Si en lugar de la sucesión numérica tenemos una sucesión de variables aleatorias, ¿cómo se puede definir el concepto de convergencia en este caso?

Veremos a continuación que puede responderse a esta pregunta de varias maneras. Consideremos entonces que tenemos una sucesión infinita de variables aleatorias  $X_1, X_2, \dots$  y un espacio de probabilidad en donde todas estas variables aleatorias están definidas. La variedad de formas en las que puede definirse la convergencia de variables aleatorias estará dada por las formas en las que se decida medir la cercanía de la sucesión con el límite a través de la medida de probabilidad.

### Convergencia puntual

Para cada  $\omega$  fijo, la sucesión  $X_1(\omega), X_2(\omega), \dots$  es una sucesión de números reales, por lo tanto podemos definir la convergencia de las variables aleatorias cuando esta sucesión numérica es convergente para cada  $\omega$  fijo. En este caso la v.a. límite se define de forma puntual:  $X(\omega) := \lim_{n \rightarrow \infty} X_n(\omega)$ . A este tipo de convergencia se le llama **convergencia puntual** y se escribe

$$X_n \rightarrow X, \quad \text{para cada } \omega \in \Omega.$$

### Convergencia casi segura

Un tipo de convergencia un tanto menos estricto que el anterior ocurre cuando se permite que la convergencia puntual se observe sobre un conjunto de probabilidad uno, es decir, se dice que la sucesión  $X_1, X_2, \dots$  converge **casi seguramente**, o **casi dondequiera**, a la variable  $X$  si para casi toda  $\omega$ ,  $X_n(\omega)$  converge a  $X(\omega)$ , en símbolos,

$$P(\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)) = 1,$$

y se escribe

$$X_n \xrightarrow{\text{c.s.}} X.$$

De este modo se permite que exista un subconjunto de  $\Omega$  en donde no se verifique la convergencia, pero tal subconjunto debe tener medida de probabilidad cero. Es claro que si una sucesión de v.a.s es convergente puntualmente, entonces es también convergente en el sentido casi seguro. El recíproco es falso.

### Convergencia en probabilidad

Otra forma aún menos restrictiva que la convergencia casi segura es la siguiente: la sucesión de variables aleatorias  $X_1, X_2, \dots$  converge **en probabilidad** a la variable aleatoria  $X$  si para cualquier  $\epsilon > 0$ ,

$$P(\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon) \rightarrow 0,$$

cuando  $n$  tiende a infinito. En este caso se escribe

$$X_n \xrightarrow{P} X.$$

### Convergencia en distribución

Finalmente el último tipo de convergencia que mencionaremos hace uso de las funciones de distribución de las variables aleatorias: se dice que la sucesión de variables  $X_1, X_2, \dots$  converge **en distribución**, o que converge **débilmente**, a la variable aleatoria  $X$  si para todo punto  $x$  en donde  $F_X(x)$  es continua se cumple que cuando  $n \rightarrow \infty$ ,

$$F_{X_n}(x) \rightarrow F_X(x).$$

Es decir, para aquellos valores reales  $x$  que cumplan la condición mencionada, debe verificarse que

$$\lim_{n \rightarrow \infty} P(X_n \leq x) = P(X \leq x).$$

En este caso se escribe

$$X_n \xrightarrow{d} X.$$

El teorema de continuidad de la función generadora de momentos que enunciamos en la página 163 se refiere precisamente a este tipo de convergencia. Existen otros tipos de convergencia para variables aleatorias pero los que hemos mencionado son suficientes para poder enunciar algunos teoremas límite importantes en probabilidad. Antes de pasar al estudio de estos resultados, responderemos a la pregunta que posiblemente se habrá hecho el lector respecto a las posibles relaciones que pudieran existir entre los tipos de convergencia de variables aleatorias:

*¿existe alguna relación entre los diferentes tipos de convergencia de variables aleatorias?*

La respuesta a esta pregunta se muestra gráficamente en la Figura 5.1, en donde un punto dentro de alguna región significa una sucesión de v.a.s que es convergente en el sentido indicado. La contención de una región en otra indica que el tipo de convergencia de la región contenida implica el tipo de convergencia de la región que contiene, así por ejemplo, la convergencia casi segura implica la convergencia en probabilidad, y ésta a su vez implica la convergencia en distribución. El diagrama establece además que las contenciones son propias, es decir, existen elementos en una región que no pertenecen a los subconjuntos contenidos, por ejemplo, existen sucesiones de v.a.s que convergen en probabilidad pero no convergen en el sentido casi seguro. El lector interesado en la demostración de las afirmaciones anteriores puede consultar textos más avanzados de probabilidad como [7].

Los teoremas límite que estudiaremos en las siguientes secciones tratan sobre la convergencia de sucesiones de variables aleatorias de la forma

$$\frac{1}{n} \sum_{i=1}^n X_i.$$

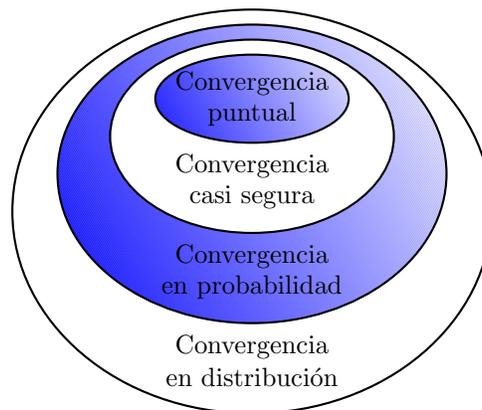


Figura 5.1

## Ejercicios

405. Para cada número natural  $n$  suponga que  $X_n$  una variable aleatoria con distribución  $\text{unif}(0, 1/n)$ . Sea  $X$  la v.a. constante cero. Demuestre que

$$X_n \xrightarrow{d} X.$$

### 5.3. La ley de los grandes números

El teorema conocido como la [ley de los grandes números](#) es un resultado muy interesante que puede observarse en la naturaleza. Constituye uno de los resultados más importantes de la teoría de la probabilidad y tiene mucha relevancia en las aplicaciones tanto teóricas como prácticas. Este teorema establece que bajo ciertas condiciones, el promedio aritmético de variables aleatorias converge a una constante cuando el número de sumandos crece a infinito. Ya desde el siglo XVI el matemático Gerolamo Cardano (1501-1576) había hecho la observación de que la precisión de las estadísticas empíricas mejoraban conforme se incrementaba el número de observaciones. Pero fue Jacobo Bernoulli, quien en 1713 y después de muchos años de trabajo logró formalizar por primera vez el enunciado del teorema y dar una demostración rigurosa para el caso de variables aleatorias dicotómicas. Debido a este gran éxito en la carrera de Jacobo Bernoulli, este resultado se conoce también como [teorema de Bernoulli](#). Sin embargo, fue Simone D. Poisson quien usó y popularizó el término *ley de los grandes números*. Otros matemáticos han contribuido notablemente a la generalización y extensión del teorema de Bernoulli, entre ellos están Chebyshev, Markov, Borel, Cantelli, Kolmogorov y Khinchin.

**Teorema 5.1 (Ley de los grandes números)** Sea  $X_1, X_2, \dots$  una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas con media finita  $\mu$ . Entonces, cuando  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow \mu.$$

en donde la convergencia se verifica en el sentido casi seguro (ley fuerte) y también en probabilidad (ley débil).

**Demostración.** (Ley débil, es decir, convergencia en probabilidad, suponiendo segundo momento finito). Recordemos que la desigualdad de Chebyshev para una variable aleatoria  $X$  con media  $\mu$  y varianza  $\sigma^2$  establece que para cualquier  $\epsilon > 0$ ,

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Así, aplicando este resultado a la variable aleatoria  $\frac{1}{n} \sum_{i=1}^n X_i$ , cuya esperanza es  $\mu$  y varianza es  $\sigma^2/n$ , tenemos que para cualquier  $\epsilon > 0$ ,

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

De modo que al hacer  $n \rightarrow \infty$  se obtiene que

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \epsilon\right) \rightarrow 0,$$

lo cual significa que

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

•

Así, sin importar la distribución de las variables aleatorias, el promedio aritmético converge a la media  $\mu$  conforme  $n$  tiende a infinito. Como se ha

mencionado, únicamente se ha presentado la demostración en el caso cuando la convergencia es en probabilidad y suponiendo adicionalmente que el segundo momento existe. Demostraciones más generales de este resultado pueden encontrarse por ejemplo en el texto de Gut [7]. La siguiente demostración alternativa de la ley débil (convergencia en probabilidad) es también bastante directa aunque tiene la desventaja de suponer la existencia de la función generadora de momentos, pero ello garantizará también la convergencia en distribución de los promedios parciales.

**Demostración.** Sea  $S_n = (X_1 + \cdots + X_n)/n$  y sea  $M_X(t)$  la f.g.m. de cualquiera de las variables  $X_i$ . Haremos uso de la expansión (2.19) de la página 161 y de la notación  $o$ -pequeña que puede consultarse en el apéndice en la página 313. La función generadora de momentos de la variable  $S_n$  es

$$\begin{aligned} M_{S_n}(t) &= E(e^{tS_n}) \\ &= E(e^{\frac{t}{n}(X_1 + \cdots + X_n)}) \\ &= (M_X(\frac{t}{n}))^n \\ &= (1 + \frac{t}{n}E(X) + o(\frac{t}{n}))^n \\ &= (1 + \frac{t}{n}E(X))^n + o(\frac{t}{n}), \end{aligned}$$

en donde se ha escrito la  $n$ -ésima potencia de un binomio en dos sumandos, uno en donde el primer término tiene exponente  $n$  y el segundo  $o(\frac{t}{n})$  desaparece pues tiene exponente cero, el segundo sumando tiene potencias positivas de  $o(\frac{t}{n})$  y todos estos términos se agrupan en una misma expresión escrita como  $o(\frac{t}{n})$ . Por lo tanto,

$$\lim_{n \rightarrow \infty} M_{S_n}(t) = e^{tE(X)},$$

en donde esta última expresión corresponde a la f.g.m. de la variable aleatoria constante igual a  $E(X)$ . Por la Proposición 2.1 de la página 163 sobre la continuidad de las funciones generadoras de momentos, tenemos que

$$S_n \xrightarrow{d} E(X).$$

Se puede demostrar que la convergencia en distribución a una constante implica la convergencia en probabilidad a la misma constante, la demostración de esta afirmación no es complicada y puede verse, por ejemplo, en [17]. Por lo tanto, también tenemos que

$$S_n \xrightarrow{p} E(X).$$

•

**Ejemplo 5.1 (Probabilidad frecuentista)** Consideremos un experimento aleatorio cualquiera y sea  $A$  un evento con probabilidad desconocida  $p$ . Nos interesa encontrar este valor de  $p$ . Suponga que se efectúan varias realizaciones sucesivas e independientes del experimento y se observa en cada ensayo la ocurrencia o no ocurrencia del evento  $A$ . Para cada entero  $i \geq 1$  defina la variable aleatoria

$$X_i = \begin{cases} 1 & \text{si ocurre el evento } A \text{ en el } i\text{-ésimo ensayo,} \\ 0 & \text{si no ocurre el evento } A \text{ en el } i\text{-ésimo ensayo.} \end{cases}$$

Entonces las variables  $X_1, X_2, \dots$  son independientes cada una con distribución  $\text{Ber}(p)$ , en donde  $p$  es la probabilidad del evento  $A$ . Por lo tanto,  $E(X_i) = p$  y  $\text{Var}(X_i) = p(1-p)$ . La ley de los grandes números asegura que la fracción de ensayos en los que se observa el evento  $A$  converge a la constante desconocida  $p$  cuando el número de ensayos crece a infinito, es decir,

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow p.$$

Esta es la definición frecuentista de la probabilidad que habíamos estudiado antes en la página 28 y que ahora hemos podido corroborar su validez con la ayuda de la teoría desarrollada a partir de los axiomas de la probabilidad de Kolmogorov y en particular de la ley de los grandes números. •

**Simulación 5.1** En este ejemplo se utiliza R para ilustrar la ley de los grandes números. Si se define la variable

$$S_n = \frac{1}{n}(X_1 + \dots + X_n),$$

**Programa en R para ilustrar la ley de los grandes números en el caso de variables aleatorias  $\text{Ber}(p)$  con  $p = 1/2$**

```
n=200
s=rep(1,n)
p=0.5
s[1]=rbinom(1,1,p)
for(i in 2:n){
  s[i]=((i-1)*s[i-1]+rbinom(1,1,p))/i
}
plot(s,type="l")      # letra ele no uno
abline(h=p)
```

Figura 5.2

entonces podemos expresar  $S_n$  en términos de  $S_{n-1}$  de la siguiente forma: para  $n \geq 2$ ,

$$S_n = \frac{1}{n}((n-1)S_{n-1} + X_n).$$

Esta expresión es muy útil para analizar numéricamente el comportamiento de  $S_n$  a lo largo del tiempo. En particular simularemos 200 lanzamientos independientes de una moneda equilibrada, es decir, se generarán 200 valores al azar de una variable aleatoria con distribución  $\text{Ber}(p)$ , con  $p = 1/2$ . El código R se muestra en la Figura 5.2. Los resultados obtenidos en R se exportaron después a  $\text{\LaTeX}$  para generar la gráfica en PSTricks que se muestra en la Figura 5.3. En esta gráfica puede apreciarse el comportamiento asintótico de los promedios  $S_n = (X_1 + \dots + X_n)/n$  conforme  $n$  crece. Los puntos graficados fueron unidos por una línea continua para una mejor visualización del comportamiento inicial oscilante y su eventual estabilización en el valor  $1/2$ . En este ejemplo hemos utilizado la distribución Bernoulli pero cualquier otra distribución puede ser usada para observar el comportamiento asintótico de la media aritmética a la media de la distribución.

•

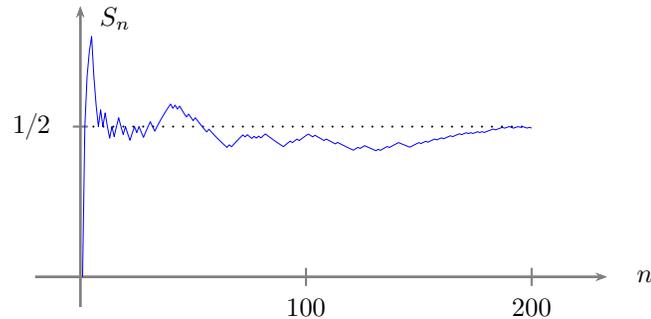


Figura 5.3: Ley de los grandes números.

## Ejercicios

### 406. Ley de los grandes números: simulación.

- a) Escoja usted una distribución de probabilidad discreta de su preferencia, especificando valores numéricos para sus parámetros. Genere  $n = 1000$  valores independientes al azar  $x_1, \dots, x_{1000}$  de esta distribución y calcule los promedios parciales

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad n = 1, 2, \dots, 1000.$$

Grafique la función  $n \mapsto \bar{x}_n$  uniendo con una línea recta sus valores. Denote por  $\mu$  la media de la distribución. Trace en la misma gráfica la función constante  $\mu$  y compruebe gráficamente que la función  $n \mapsto \bar{x}_n$  oscila y se aproxima al valor  $\mu$ . Esta es una comprobación experimental de la ley de los grandes números.

- b) Haga lo mismo que en el inciso anterior ahora con una distribución continua de su preferencia.

407. **Estimaciones.** Sea  $x_1, x_2, \dots, x_n$  una colección finita de observaciones de una v.a. con distribución normal con media  $\mu$  y varianza  $\sigma^2$  desconocidas. Con base en la ley de los grandes números proporcione expresiones en función de  $x_1, x_2, \dots, x_n$  que puedan servir para estimar los valores de  $\mu$  y  $\sigma^2$ .

408. **Convergencia de la media geométrica.** Sea  $X_1, X_2, \dots$  una sucesión de v.a.s positivas, independientes, idénticamente distribuidas y tales que  $E(\ln X) = \mu < \infty$ . Demuestre que

$$\sqrt[n]{X_1 X_2 \cdots X_n} \xrightarrow{c.s.} e^\mu.$$

409. **Teorema de equipartición asintótica.** La entropía en base 2 de una v.a. discreta  $X$  con función de probabilidad  $p(x)$  se define como

$$H(X) := \sum_x p(x) \log_2 p(x).$$

Demuestre que si  $X_1, X_2, \dots$  es una sucesión de v.a.s discretas, independientes, idénticamente distribuidas, con función de probabilidad  $p(x)$  y con entropía finita, entonces

$$-\frac{1}{n} \log_2 p(X_1, X_2, \dots, X_n) \xrightarrow{c.s.} H(X).$$

En la teoría de la información a este resultado se le conoce como el teorema de equipartición asintótica.

410. **Una versión de la ley débil de los grandes números.** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias independientes no necesariamente idénticamente distribuidas pero con media común  $\mu$  y varianzas acotadas, es decir,  $\text{Var}(X_i) \leq c$ , en donde  $c$  es una constante. Use la desigualdad de Chebyshev para demostrar que

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

411. **Otra versión de la ley débil de los grandes números.** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias no necesariamente idénticamente distribuidas ni independientes pero con media común  $\mu$  y varianzas tales que  $\text{Var}(\frac{1}{n} \sum_{i=1}^n X_i) \rightarrow 0$  cuando  $n \rightarrow \infty$ . Use la desigualdad de Chebyshev para demostrar que

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

## 5.4. El teorema central del límite

Este teorema es muy importante y tiene una amplia gama de aplicaciones para simplificar el cálculo de ciertas probabilidades y aproximar algunas distribuciones.

**Teorema 5.2 (Teorema central del límite)** Sea  $X_1, X_2, \dots$  una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas, con media  $\mu$  y varianza finita  $\sigma^2$ . Entonces la función de distribución de la variable aleatoria

$$Z_n = \frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n\sigma^2}}$$

tiende a la función de distribución normal estándar cuando  $n$  tiende a infinito.

**Demostración.** Por simplicidad supondremos que, a diferencia del enunciado del teorema, las variables aleatorias tienen todos sus momentos finitos de tal manera que su f.g.m. es finita y tiene la expresión de la serie de potencias (2.19) de la página 161. Nuevamente haremos uso de la notación  $o$ -pequeña. Podemos calcular la f.g.m. de la variable  $Z_n$  de la siguiente manera:

$$\begin{aligned} M_{Z_n}(t) &= E(e^{tZ_n}) \\ &= E\left(e^{\frac{t}{\sqrt{n}}\left(\frac{X_1-\mu}{\sigma} + \dots + \frac{X_n-\mu}{\sigma}\right)}\right) \\ &= \left(M_{\frac{X-\mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{t}{\sqrt{n}}E\left(\frac{X-\mu}{\sigma}\right) + \frac{t^2}{2n}E\left(\frac{X-\mu}{\sigma}\right)^2 + o\left(\frac{t^2}{n}\right)\right)^n \\ &= \left(1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \\ &= \left(1 + \frac{t^2}{2n}\right)^n + o\left(\frac{t^2}{n}\right), \end{aligned}$$

en donde se ha escrito la  $n$ -ésima potencia de un binomio en dos suman-

dos, uno en donde el primer término tiene exponente  $n$  y el segundo  $o(\frac{t^2}{n})$  desaparece pues tiene exponente cero, el segundo sumando tiene potencias positivas de  $o(\frac{t^2}{n})$  y todos estos términos se agrupan en una misma expresión escrita como  $o(\frac{t^2}{n})$ . Por lo tanto,

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2},$$

en donde esta última expresión corresponde a la f.g.m. de una variable aleatoria con distribución  $N(0, 1)$ . Por la continuidad de las funciones generadoras de momentos,

$$Z_n \xrightarrow{d} N(0, 1).$$

En términos matemáticos este resultado establece que para cualquier número real  $x$ ,

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x),$$

sin importar la distribución de las variables  $X_1, X_2, \dots$ , así es que éstas pueden tener distribución Bernoulli, binomial, exponencial, gama, etc., en general pueden ser discretas o continuas, y este resultado sorprendente asegura que la variable  $Z_n$  tiene una distribución aproximada normal estándar para valores grandes de  $n$ . Una forma de expresar el teorema central del límite de manera informal es a través de la siguiente afirmación: para cualesquiera números reales  $a < b$ ,

$$P(a < \frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n\sigma^2}} < b) \approx \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Esto nos permitirá aproximar probabilidades de eventos que involucran sumas de variables aleatorias en términos de probabilidades de la distribución normal estándar. Observe que dividiendo el numerador y denominador entre  $n$ , y definiendo  $\bar{X} = (X_1 + \dots + X_n)/n$ , la variable  $Z_n$  puede escribirse de la siguiente forma

$$Z_n = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}. \quad (5.3)$$

Es interesante observar también que la ley de los grandes números asegura que el numerador de (5.3) converge a cero conforme  $n$  tiende a infinito, sin

embargo el denominador de esta expresión también converge a cero y estos límites ocurren de tal manera que este cociente no es constante sino una variable aleatoria con distribución normal estándar.

Una de las primeras versiones demostradas del teorema central del límite es aquella en donde las variables aleatorias tienen distribución Bernoulli. Enunciamos a continuación este resultado en el contexto de ocurrencias o no ocurrencias de un evento en una sucesión de ensayos independientes de un experimento aleatorio cualquiera.

### Teorema 5.3 (Teorema de De Moivre-Laplace)

Suponga que se tiene una sucesión infinita de ensayos independientes de un experimento aleatorio. Sea  $A$  un evento de este experimento aleatorio con probabilidad de ocurrencia  $p > 0$ . Sea  $n_A$  el número de ocurrencias del evento de interés en los primeros  $n$  ensayos del experimento. Entonces para cualesquiera números reales  $a < b$ ,

$$\lim_{n \rightarrow \infty} P\left(a < \frac{n_A/n - p}{\sqrt{p(1-p)/n}} < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

**Simulación 5.2** En el código que aparece en la Figura 5.4 se muestra la forma en la que puede usarse R para comprobar mediante simulación el teorema central del límite. Como en el enunciado del teorema, el parámetro  $n$  corresponde al número de sumando  $X_1 + \dots + X_n$ . El parámetro  $k = 1000$  se usa para generar  $k$  valores al azar de la variable aleatoria suma  $X_1 + \dots + X_n$  y de esa forma aproximar su función de distribución. En estas simulaciones se ha utilizado la distribución  $\text{Ber}(p)$  con  $p = 0.7$  para los sumandos, lo cual puede modificarse con facilidad. Los resultados aparecen en la Figura 5.5 para  $n = 5, 50, 100, 200, 500$  y  $1000$ . Puede apreciarse con claridad la forma sorprendente en la que la función de distribución de la variable

$$Z_n = \frac{(X_1 + \dots + X_n) - n\mu}{\sqrt{n\sigma^2}}$$

se hace cada vez más parecida a la función de distribución normal estándar. Para obtener las gráficas mostradas en la Figura 5.5 se obtuvieron primero

los datos en R usando el ambiente gráfico RStudio, se trasladaron después estos datos a  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  para su graficación a través del paquete PSTricks. A fin de que aparecieran únicamente líneas horizontales y verticales en la función de distribución aproximante, se llevó a cabo un proceso de interpolación en los subintervalos en donde aparecían líneas inclinadas. Las últimas dos gráficas requirieron un refinamiento en el número de puntos a graficar. Para este mismo ejemplo se muestran en la Figura 5.6 varios histogramas que paulatinamente adquieren la forma de la función de densidad normal estándar conforme el parámetro  $n$  crece. En este tipo de gráficas y para hacer comparaciones entre dos histogramas, es importante escoger de manera adecuada el tamaño de la base de los rectángulos. •

**Ejemplo 5.2** Se lanza una dado repetidas veces y sean  $X_1, X_2, \dots$  los resultados de estos lanzamientos. Es razonable suponer que estas variables aleatorias son independientes y con idéntica distribución uniforme en el conjunto  $\{1, 2, 3, 4, 5, 6\}$ . En particular, la esperanza es  $\mu = 3.5$  y la varianza es  $\sigma^2 = 2.91\bar{6}$ . Por la ley de los grandes números, sabemos que el promedio parcial  $\bar{X} = (X_1 + \dots + X_n)/n$  se aproxima a la media 3.5 conforme  $n$  crece. ¿Cuántas veces debe lanzarse el dado de tal forma que  $\bar{X}$  se encuentre entre 3 y 4 con una probabilidad de 0.99?

Solución. Se busca el valor de  $n$  tal que

$$P(3 \leq \bar{X} \leq 4) = 0.99.$$

Restando en cada lado de las desigualdades la media  $\mu$  y dividiendo entre  $\sqrt{\sigma^2/n}$ , la igualdad anterior es equivalente a la ecuación

$$P\left(\frac{3 - 3.5}{\sqrt{\sigma^2/n}} \leq \frac{\bar{X} - 3.5}{\sqrt{\sigma^2/n}} \leq \frac{4 - 3.5}{\sqrt{\sigma^2/n}}\right) = 0.99.$$

Por el teorema central del límite, la probabilidad indicada es aproximadamente igual a  $P(-0.5/\sqrt{\sigma^2/n} \leq Z \leq 0.5/\sqrt{\sigma^2/n})$ , en donde  $Z$  es una variable aleatoria con distribución normal estándar. Es decir, tenemos ahora la ecuación de aproximación

$$\Phi\left(\frac{0.5}{\sqrt{\sigma^2/n}}\right) - \Phi\left(\frac{-0.5}{\sqrt{\sigma^2/n}}\right) = 0.99.$$

**Programa en R para ilustrar el teorema central del límite  
en el caso de variables aleatorias  $\text{Ber}(p)$  con  $p = 0.7$** 

```
# Número de valores simulados de la v.a. suma s[1],...,s[k]:
k=1000
s=rep(0,k)
# Número de sumandos x[1],...,x[n]:
n=5
x=rep(0,n)
# Parámetro(s):
p=0.7
# Generación al azar de n sumandos x[i] y k sumas s[i]:
for (i in 1:k){
  x=rbinom(n,1,p)
  s[i]=sum(x)
}
# Cálculo de media, varianza y estandarización
media=n*p
var=n*p*(1-p)
s=(s-media)/sqrt(var)
# Graficación de la función de densidad
par(mfrow=c(1,2))
curve(dnorm(x),from=-3,to=3,ylim=c(0,0.6),ylab="F. de
densidad",lwd=2,col="blue")
hist(s,freq=FALSE,breaks=50,add=T,xlim=c(-3,3),ylim=c(0,5))
# Graficación de la función de distribución
distempirica<-ecdf(s)
curve(distempirica,from=-3,to=3,cex=0.1,ylab="F. de dist.")
curve(pnorm(x),from=-3,to=3,add=TRUE,col="blue")
```

Figura 5.4

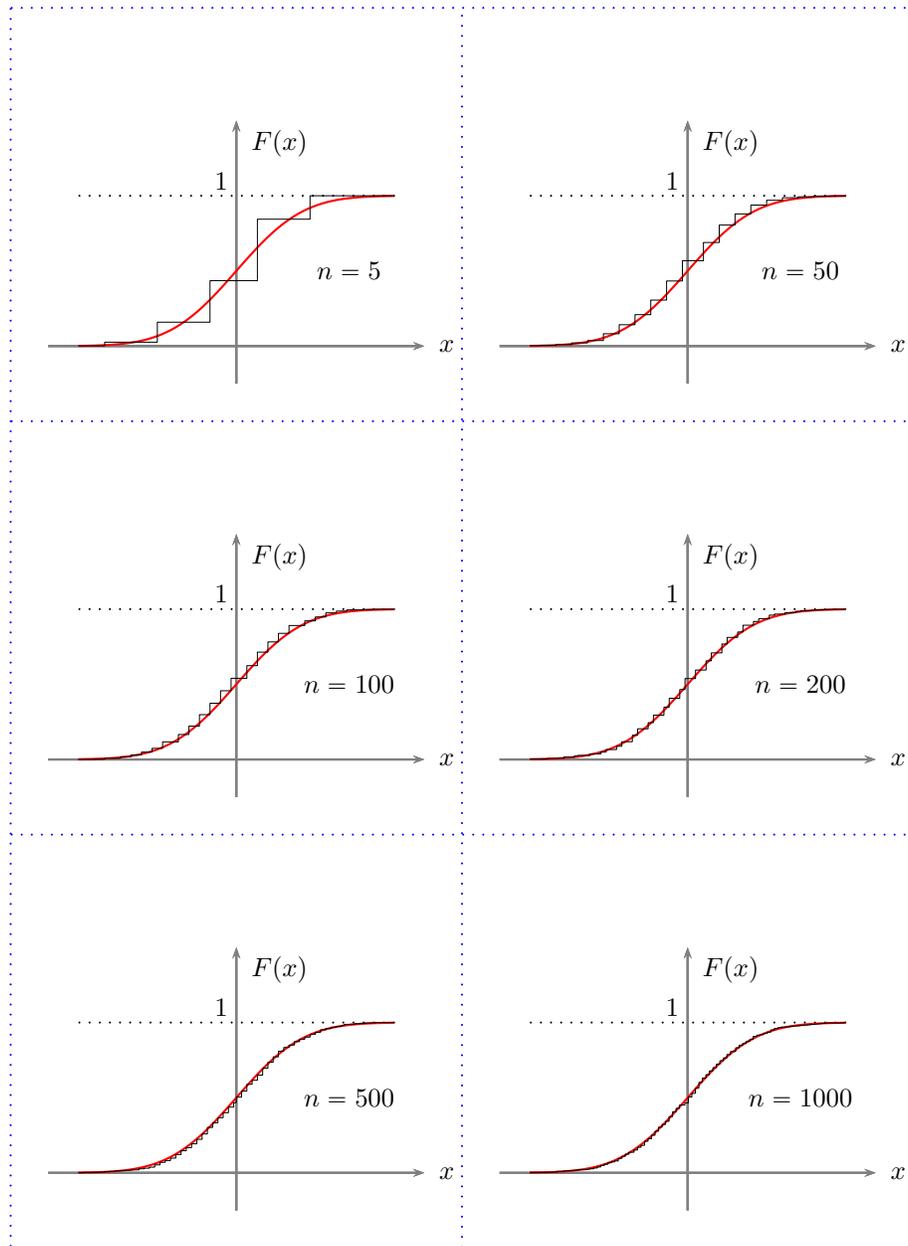


Figura 5.5: Teorema central del límite.

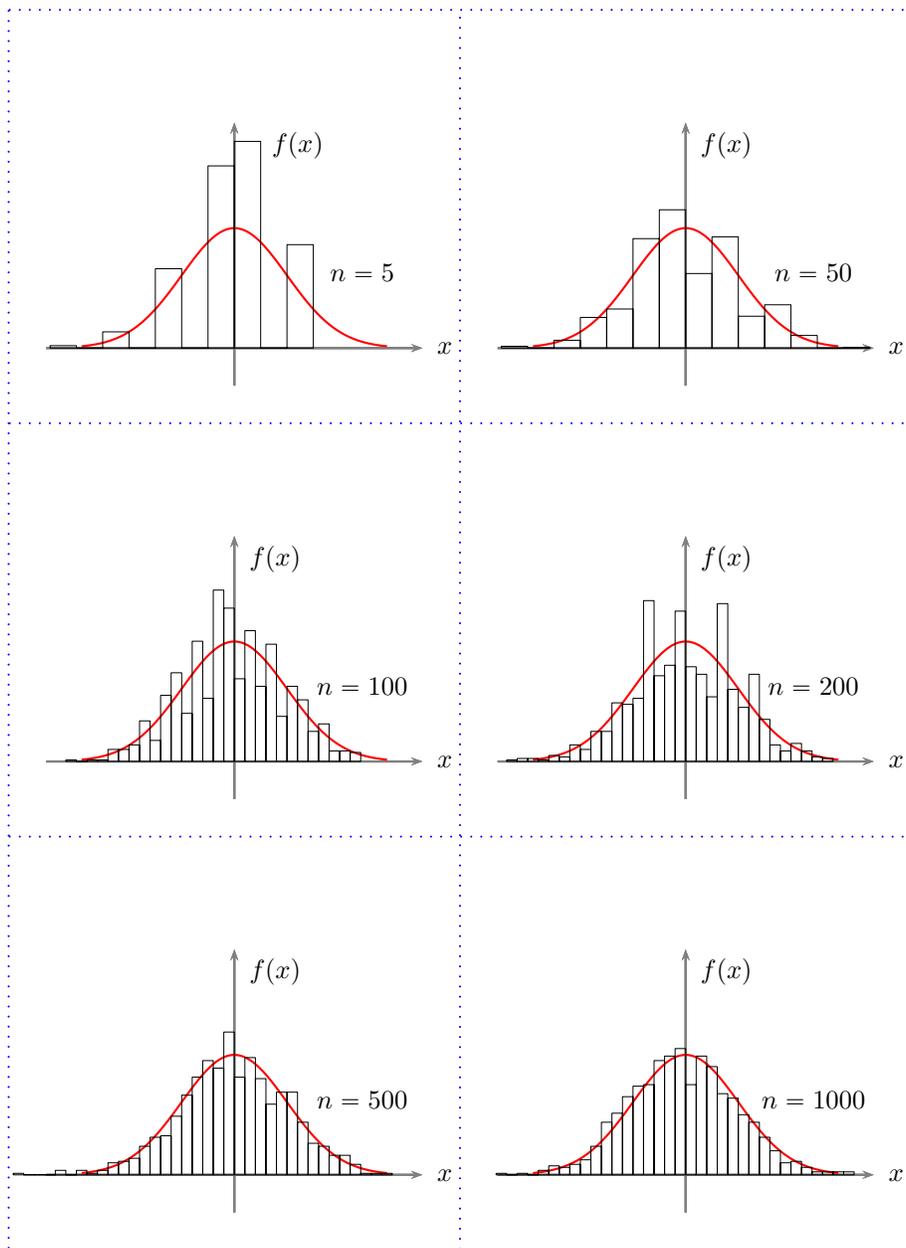


Figura 5.6: Teorema central del límite.

De tablas de la distribución normal estándar puede verificarse que el valor de  $x$  tal que  $\Phi(x) - \Phi(-x) = 0.99$  es  $x = 2.58$ . De este modo se tiene que  $0.5/\sqrt{\sigma^2/n} = 2.58$ , de donde se obtiene  $n = 226.5$ . •

**Ejemplo 5.3** Se desea diseñar un estacionamiento de coches para un conjunto de 200 departamentos que se encuentran en construcción. Suponga que para cada departamento, el número de automóviles será de 0, 1 o 2, con probabilidades 0.1, 0.6 y 0.3, respectivamente. Se desea que con una certeza del 95 % haya espacio disponible para todos los coches cuando los departamentos se vendan. ¿Cuántos espacios de estacionamiento deben construirse? Solución. Sean  $X_1, \dots, X_{200}$  las variables aleatorias que denotan el número de automóviles que poseen los futuros dueños de los departamentos. Podemos suponer que estas variables aleatorias discretas son independientes unas de otras y todas ellas tienen la misma distribución de probabilidad:  $P(X = 0) = 0.1$ ,  $P(X = 1) = 0.6$ , y  $P(X = 2) = 0.3$ . De esta forma la variable aleatoria suma  $X_1 + \dots + X_{200}$  denota el total de automóviles que habrá en el complejo de departamentos. Se desconoce la distribución de esta variable aleatoria, sin embargo se desea encontrar el valor de  $n$  tal que  $P(X_1 + \dots + X_{200} \leq n) = 0.95$ . Haremos uso del teorema central del límite para resolver este problema, y para ello se necesita calcular la esperanza y varianza de  $X$ . Puede comprobarse que  $E(X) = 1.2$  y  $\text{Var}(X) = 0.36$ , cantidades que denotaremos por  $\mu$  y  $\sigma^2$  respectivamente. La ecuación planteada es entonces

$$P(X_1 + \dots + X_{200} \leq n) = 0.95,$$

en donde la incógnita es el valor de  $n$ . Restando en ambos lados de la desigualdad  $200\mu$  y dividiendo entre  $\sqrt{200\sigma^2}$ , la ecuación anterior es equivalente a

$$P\left(\frac{X_1 + \dots + X_{200} - 200\mu}{\sqrt{200\sigma^2}} \leq \frac{n - 200\mu}{\sqrt{200\sigma^2}}\right) = 0.95. \quad (5.4)$$

Por el teorema central del límite, la probabilidad indicada es aproximada a  $\Phi((n - 200\mu)/\sqrt{200\sigma^2})$ . De este modo tenemos ahora la ecuación

$$\Phi\left(\frac{n - 200\mu}{\sqrt{200\sigma^2}}\right) = 0.95.$$

Observe que la variable aleatoria que aparece en la ecuación (5.4) y cuya distribución de probabilidad se desconoce y en general es difícil conocer, se ha aproximado por una variable aleatoria normal estándar, y allí radica la utilidad del teorema central del límite. De la tabla de la distribución normal podemos ahora verificar que el valor de  $x$  tal que  $\Phi(x) = 0.95$  es  $x = 1.65$ . De este modo se llega a la igualdad  $(n - 200\mu)/\sqrt{200\sigma^2} = 1.65$ , de donde se obtiene que  $n = 253.99$ . Es decir, el tamaño del estacionamiento debe ser de aproximadamente 254 lugares. •

## Ejercicios

### 412. Teorema central del límite: simulación.

- a) Escoja usted una distribución de probabilidad discreta de su preferencia especificando valores numéricos para sus parámetros. Denote por  $\mu$  a la media de la distribución y sea  $\sigma^2$  su varianza. Lleve a cabo las indicaciones de los siguientes incisos para:

$$n = 20, 40, 60, 80, 100.$$

$$N = 50, 100.$$

Esta es una comprobación del teorema central del límite.

- 1) Genere  $n$  valores independientes al azar  $x_1, \dots, x_n$  y calcule el promedio

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

- 2) Repita  $N$  veces el inciso anterior calculando los promedios centrados

$$\underbrace{\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}, \dots, \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}}_N.$$

- 3) Elabore un histograma de estos  $N$  valores trazando en la misma gráfica la función de densidad  $N(0, 1)$ . Utilice un tamaño adecuado para la base de los rectángulos.

- 4) Elabore una gráfica de la frecuencia relativa acumulada (función de distribución empírica) de los  $N$  valores uniendo los puntos con una línea recta y en la misma gráfica dibuje la función de distribución  $N(0, 1)$ .
- b) Haga lo mismo que en el inciso anterior ahora con una distribución continua de su preferencia.
413. Sea  $A$  un evento de un experimento aleatorio cuya probabilidad es  $p \in (0, 1)$ . Suponga que se efectúan  $n$  ensayos independientes del experimento y denote por  $n_A$  el número de veces que se observó la ocurrencia del evento  $A$ . Demuestre que para  $n$  suficientemente grande y para cualquier  $\epsilon > 0$ ,

$$P\left(p - \epsilon \leq \frac{n_A}{n} \leq p + \epsilon\right) \approx \Phi\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right).$$

Sugerencia: Use el teorema de Moivre-Laplace.

414. La probabilidad de que un componente electrónico falle durante ciertas pruebas de control de calidad es 0.05. Use el teorema de Moivre-Laplace para encontrar una aproximación de la probabilidad de que al probar 100 componentes, el número de fallas sea:
- al menos 5.
  - mayor a 5.
  - entre 5 y 10.
415. La probabilidad de un cierto evento en un ensayo de un experimento aleatorio es 0.7. ¿Cuál es la probabilidad de que este evento aparezca en la mayoría de una sucesión de 200 ensayos independientes?
416. La probabilidad de ocurrencia de un evento en un ensayo es 0.3. ¿Cuál es la probabilidad de que la frecuencia relativa de este evento en 100 ensayos se encuentre entre 0.2 y 0.5?
417. Sea  $X$  con distribución  $\text{bin}(n, p)$  y sea  $Z$  con distribución normal estándar. Demuestre que cuando  $n \rightarrow \infty$ ,

$$\frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z.$$

418. Sea  $X$  con distribución  $\chi^2(n)$  y sea  $Z$  con distribución normal estándar. Demuestre que cuando  $n \rightarrow \infty$ ,

$$\frac{X - n}{\sqrt{2n}} \xrightarrow{d} Z.$$

419. Sean  $X_1, \dots, X_n$  v.a.s independientes cada una de ellas con distribución Poisson( $\lambda$ ). Encuentre una aproximación para las siguientes probabilidades en términos de la función de distribución  $\Phi(x)$  de la distribución normal estándar:

- a)  $P(a < X_1 + \dots + X_n < b)$ .
- b)  $P(X_1 + \dots + X_n \geq n\lambda)$ .

## Jacobo Bernoulli

J. Bernoulli (Suiza, 1654–1705) fue uno de los matemáticos más sobresalientes de la familia Bernoulli. Sus padres tuvieron una posición económica cómoda, Jacobo Bernoulli fue el primero de los hijos y por instrucciones de sus padres estudió filosofía y teología. Se graduó de la Universidad de Basilea en Suiza con el grado de maestro en filosofía en 1671 y obtuvo también una *licenciate* en teología en 1676. Durante sus estudios universitarios y contrario al deseo de sus padres, Jacobo



J. Bernoulli

Bernoulli también estudió matemáticas y astronomía. Con el paso del tiempo fue claro que el verdadero amor de Jacobo Bernoulli no fue la teología ni la filosofía, sino las matemáticas y la física teórica, pues en estas disciplinas impartió clases y desarrollo trabajos científicos de primer nivel. Cronológicamente Jacobo Bernoulli fue el primer matemático de la familia Bernoulli y posiblemente haya sido de cierta influencia para que algunos de los siguientes miembros de la familia decidieran seguir una carrera científica.

Después de graduarse de la Universidad de Basilea, se trasladó a Ginebra en donde trabajó como tutor. En los siguientes años viajó a Francia, Holanda e Inglaterra en donde conoció y mantuvo comunicación con renombrados matemáticos y científicos de estos países. En 1683 regresó a Suiza y empezó a trabajar en la Universidad de Basilea habiendo ya empezado a publicar sus trabajos un año antes. En dicha universidad fue contratado como profesor de matemáticas en 1687. Produjo trabajos de mucha trascendencia en las áreas del cálculo infinitesimal, el álgebra, la teoría de series, el cálculo de variaciones, la mecánica y particularmente la teoría de la probabilidad. En 1713, ocho años después de la muerte de Jacobo Bernoulli, se publica una de sus obras más preciadas: *Ars Conjectandi* (El arte de conjeturar). A este trabajo de Bernoulli se le considera como una obra fundacional del cálculo combinatorio y la teoría de la probabilidad. En este trabajo aparece por primera vez una demostración rigurosa de la ley de los grandes números en el caso cuando las variables aleatorias tienen distribución Bernoulli. A este importante resultado se le conoce ahora justamente como teorema de

Bernoulli.

Posiblemente uno de los sucesos más relevantes en la vida de Jacobo Bernoulli fue la penosa rivalidad que mantuvo con su hermano menor Johann Bernoulli (1667-1748). Era el año de 1687 cuando Johann le pidió a su hermano Jacobo que le enseñara matemáticas. Los dos hermanos empezaron así a estudiar juntos el difícil cálculo diferencial e integral de Leibnitz y otros trabajos relacionados. Ambos empezaron a resolver problemas importantes en el área pero eventualmente surgió en ellos la rivalidad por buscar cada uno un mayor reconocimiento que el otro. Jacobo, como hermano mayor y mentor, sentía que Johann debía sus éxitos a él. Johann, por su parte y posiblemente mejor matemático que Jacobo, sentía que sus méritos eran propios. Esta rivalidad persistió hasta la muerte de Jacobo Bernoulli, acaecida el 16 de agosto de 1705 a la temprana edad de 50 años. El puesto de profesor de matemáticas que Jacobo tenía en la Universidad de Basilea lo ocupó su hermano Johann.

Jacobo Bernoulli tuvo fascinación por la espiral logarítmica. Ésta es una curva que aparece frecuentemente en la naturaleza y está dada por la ecuación en coordenadas polares  $\theta = \log_b(r/a)$ . Jacobo Bernoulli pensaba que sus propiedades eran casi mágicas y que representaba un símbolo de permanencia eterna y de constante restauración exacta al ser perfecto. Por ello es que dejó instrucciones para que

en su lápida fuera grabada la inscripción en latín *Eadem Mutata Resurgo*, que significa “Resurgiré nuevamente aunque cambiado”.

En honor a la enorme contribución a la ciencia por parte de la extensa familia Bernoulli y a sugerencia del matemático y estadístico Jerzy Neyman, la agrupación internacional más importante en probabilidad y estadística lleva el nombre de [Bernoulli Society](#) y en su logo aparece la espiral logarítmica y la frase del epitafio de Jacobo Bernoulli.



Fuente: Archivo MacTutor, Universidad de St. Andrews [22].



# Apéndice A

## A.1. Notación

- $\mathbb{N}$  Conjunto de números naturales  $1, 2, 3, \dots$   
 $\mathbb{Z}$  Conjunto de números enteros  $0, \pm 1, \pm 2, \pm 3, \dots$   
 $\mathbb{Q}$  Conjunto de números racionales  $a/b$  en donde  $a, b \in \mathbb{Z}$  con  $b \neq 0$ .  
 $\mathbb{R}$  Conjunto de números reales.

## A.2. El alfabeto griego

$A \alpha$	alfa	$I \iota$	iota	$P \rho, \varrho$	ro
$B \beta$	beta	$K \kappa$	kapa	$\Sigma \sigma, \varsigma$	sigma
$\Gamma \gamma$	gama	$\Lambda \lambda$	lambda	$T \tau$	tau
$\Delta \delta$	delta	$M \mu$	mu	$\Upsilon \upsilon$	upsilon
$E \epsilon, \varepsilon$	epsilon	$N \nu$	nu	$\Phi \phi, \varphi$	fi
$Z \zeta$	zeta	$\Xi \xi$	xi	$X \chi$	ji
$H \eta$	eta	$O \omicron$	omicron	$\Psi \psi$	psi
$\Theta \theta, \vartheta$	teta	$\Pi \pi$	pi	$\Omega \omega$	omega

### A.3. Exponentes y logaritmos

#### Exponentes

- a)  $x^1 = x$ .
- b)  $x^0 = 1, \quad x \neq 0$ .
- c)  $x^{-1} = \frac{1}{x}, \quad x \neq 0$ .
- d)  $x^n x^m = x^{n+m}$ .
- e)  $\frac{x^n}{x^m} = x^{n-m}$ .
- f)  $(x^n)^m = x^{nm}$ .
- g)  $(xy)^n = x^n y^n$ .
- h)  $\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$ .
- i)  $x^{-n} = \frac{1}{x^n}, \quad x \neq 0$ .
- j)  $x^{m/n} = \sqrt[n]{x^m}$ .

#### Logaritmos

- a)  $\log ab = \log a + \log b$ .
- b)  $\log \frac{a}{b} = \log a - \log b$ .
- c)  $\log a^n = n \log a$ .
- d)  $\log \sqrt[n]{a} = \frac{1}{n} \log a$ .
- e)  $\log 1 = 0$ .
- f)  $\log_a a = 1$ .

## A.4. Fórmulas para sumas

$$\text{a) } \sum_{k=m}^n x_k = x_m + x_{m+1} + \cdots + x_n, \quad m \leq n.$$

$$\text{b) } \sum_{k=1}^n c = nc, \quad c \text{ constante.}$$

$$\text{c) } \sum_{k=1}^n k = \frac{n(n+1)}{2}.$$

$$\text{d) } \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

$$\text{e) } \sum_{k=1}^n k^3 = \left[ \frac{n(n+1)(2n+1)}{2} \right]^2.$$

$$\text{f) } \sum_{k=m}^n a^k = \frac{a^m - a^{n+1}}{1-a}, \quad a \neq 1.$$

$$\text{g) } \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, \quad x \in \mathbb{R}.$$

$$\text{h) } \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} = (a+b)^n, \quad a, b \in \mathbb{R}, n \in \mathbb{N}.$$

$$\text{i) } \sum_{k=1}^{\infty} \frac{1}{k} \text{ es divergente.}$$

$$\text{j) } \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} \quad (\text{Euler}).$$

## A.5. Fórmulas de derivación e integración

### Derivadas

- a)  $\frac{d}{dx} c = 0$ ,  $c$  constante.
- b)  $\frac{d}{dx} x = 1$ .
- c)  $\frac{d}{dx} x^n = nx^{n-1}$ .
- d)  $\frac{d}{dx} e^x = e^x$ .
- e)  $\frac{d}{dx} \ln x = \frac{1}{x}$ .
- f)  $\frac{d}{dx} \operatorname{sen} x = \cos x$ .
- g)  $\frac{d}{dx} \cos x = -\operatorname{sen} x$ .
- h)  $\frac{d}{dx} [f(x) \pm g(x)] = f'(x) \pm g'(x)$ .
- i)  $\frac{d}{dx} [f(x)g(x)] = f(x)g'(x) + f'(x)g(x)$ .
- j)  $\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}$ .
- k)  $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$ . (Regla de la cadena)

### Integrales

- a)  $\int df(x) = \int f'(x) dx = f(x) + c$ .

- b)  $\int c \, dx = c \int dx$ ,  $c$  constante.
- c)  $\int x^n \, dx = \frac{x^{n+1}}{n+1} + c$ ,  $n \neq -1$ .
- d)  $\int \frac{dx}{x} = \ln x + c$ .
- e)  $\int e^{ax} \, dx = \frac{1}{a} e^{ax} + c$ .
- f)  $\int \ln x \, dx = x \ln x - x + c$ .
- g)  $\int \operatorname{sen} x \, dx = -\cos x + c$ .
- h)  $\int \cos x \, dx = \operatorname{sen} x + c$ .
- i)  $\int u \, dv = uv - \int v \, du$ . (Integración por partes)

## A.6. El lema de Abel

Sea  $a_0, a_2, \dots$  una sucesión de números reales o complejos tal que  $\sum_{n=0}^{\infty} a_n$  es convergente. Entonces la función real  $G(t) = \sum_{n=0}^{\infty} a_n t^n$  es continua por la izquierda en  $t = 1$ , es decir,

$$\lim_{t \nearrow 1} G(t) = \sum_{n=0}^{\infty} a_n.$$

## A.7. Notación $o$ -pequeña

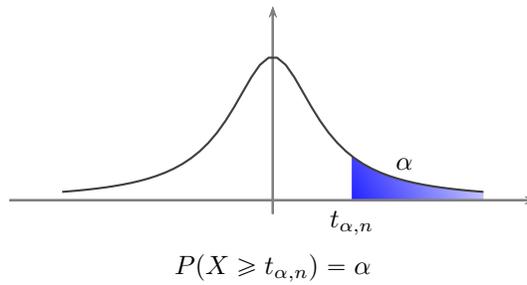
Se dice que una función  $f(x)$  definida en un intervalo no trivial alrededor del cero es  $o$ -pequeña de  $x$  cuando  $x \rightarrow 0$  si

$$\lim_{x \rightarrow 0} \frac{f(x)}{x} = 0.$$

Esto significa que la función  $f(x)$  tiende a cero cuando  $x \rightarrow 0$  más rápidamente de lo que lo hace  $x \rightarrow 0$ . Las funciones  $f(x) = x^k$  con  $k \geq 2$  son ejemplos de funciones  $o(x)$  cuando  $x \rightarrow 0$ , y se escribe  $f(x) = o(x)$  cuando  $x \rightarrow 0$ .

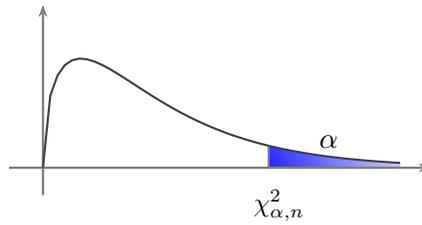


### A.9. Tabla de la distribución $t(n)$



$n \setminus \alpha$	0.005	0.01	0.025	0.05	0.1
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.474	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
$\infty$	2.576	2.326	1.960	1.645	1.282

**A.10. Tabla de la distribución  $\chi^2(n)$**



$$P(X \geq \chi^2_{\alpha, n}) = \alpha$$

$n \setminus \alpha$	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.0	0.0	0.0	0.0	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17



# Bibliografía

- [1] Blake I. F. *An introduction to applied probability*. John Wiley & Sons, 1979.
- [2] Blomm G., Holst L., Sandell D. *Problems and snapshots from the world of probability*. Springer-Verlag, 1994.
- [3] Esparza Núñez S. *Elementos de probabilidad*. IPN, 1985.
- [4] Garza T. *Elementos de cálculo de probabilidades*. UNAM, 1983.
- [5] Granville W. A. *Cálculo diferencial e integral*. Limusa, 1982.
- [6] Godfrey M. G., Roebuck E. M., Sherlock A. J. *Concise statistics*. Edward Arnold, 1988.
- [7] Gut, A. *Probability: a graduate course*. Springer, 2005.
- [8] Hoel P. G., Port S. C., Stone C. J. *Introduction to probability theory*. Houghton Mifflin, 1971.
- [9] Kolmogorov A. N. *Foundations of the theory of probability*. Chelsea Publishing Company, 1950.
- [10] Mood A. M., Graybill F. A., Boes D. C. *Introduction to the theory of statistics*. McGraw Hill, 1983.
- [11] Miller I., Miller M. *John E. Freund's mathematical statistics*. Prentice Hall, 1999.
- [12] Olofsson P. *Probability, statistics, and stochastic processes*. Wiley, 2005.

- 
- [13] Perero M. *Historia e historias de las matemáticas*. Grupo Editorial Iberoamérica, 1994.
- [14] Ross S. M. *A first course in probability*. Prentice Hall, 2009.
- [15] Ross S. M. *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2009.
- [16] Ross S. M. *Simulation*. Academic Press, 2006.
- [17] Rincón L. *Curso intermedio de probabilidad*. Facultad de Ciencias, UNAM, 2007.
- [18] Rincón L. *Curso elemental de probabilidad y estadística*. Serie Textos, SMM, 2013.
- [19] Székely G. J. *Paradoxes in probability theory and mathematical statistics*. D. Reidel Publishing Company, 2001.
- [20] <http://www.wikipedia.org>
- [21] <http://www.r-project.org/>
- [22] <http://www-history.mcs.st-and.ac.uk/>

# Índice analítico

- $\sigma$ -álgebra, 42
  - de Borel, 45
- Álgebra, 42
- Bernoulli
  - ensayo, 169
- Bernoulli, J., 306
- Borelianos, 45
- Coefficiente
  - binomial, 54
  - de correlación, 277
    - propiedades, 279
  - multinomial, 55
- Combinaciones, 53
- Conjunto
  - s ajenos, 10
  - s operaciones, 8
  - potencia, 11
- Conjuntos
  - Borel medibles, 45
  - de Borel, 45
- Convergencia
  - casi donde quiera, 285
  - casi segura, 285
  - de variables aleatorias, 284
  - débil, 286
  - en distribución, 286
    - en probabilidad, 286
    - puntual, 285
- Covarianza, 274
  - propiedades, 276
- Cuantiles, 152
- Cuartiles, 153
- De Morgan
  - leyes de, 10
- Densidad
  - conjunta, 246
- Desigualdad
  - de Chebyshev, 281
  - de Markov, 282
- Desviación estándar, 143
- Diagrama de árbol, 12
- Diferencia simétrica, 10, 15
- Distribución
  - $\chi^2$ 
    - f.g.m., 233
    - momentos, 233
  - Bernoulli, 169
  - beta, 215
  - bin. neg.
    - f.g.m., 189
    - f.g.p., 189
  - binomial, 174
    - f.g.m., 179

- f.g.p., 179
  - binomial negativa, 185
  - condicional, 268
  - conjunta, 253
  - de una v.a., 119
  - Erlang, 212
  - exponencial, 206
    - discretización, 210
    - f.g.m., 209
    - mediana, 209
    - momentos, 209
    - pérdida de memoria, 210
  - F, 237
  - gama, 210
    - f.g.m., 214
  - geométrica, 181
    - f.g.m., 184
    - f.g.p., 184
  - hipergeométrica, 189
  - ji-cuadrada, 229
  - marginal, 262
  - mixta, 122
  - normal, 221
    - f.g.m., 227
  - normal estándar, 222
  - Poisson, 193
    - f.g.m., 201
    - f.g.p., 199
  - t, 233
  - unif. cont.
    - f.g.m., 205
  - uniforme continua, 202
  - uniforme discreta, 165
  - Weibull, 218
- Distribución de prob.
- de una v.a., 119
- Distribuciones
- condicionales, 272
- Ensayo Bernoulli, 169
- Espacio
- de probabilidad, 46
  - equiprobable, 18
  - muestral, 5
- Esperanza
- de una fn. de una v.a., 133
  - de una v.a., 130
  - propiedades, 136
- Estandarización, 223
- Evento, 5
- s ajenos, 10
  - compuesto, 6
  - simple, 6
- Experimento
- aleatorio, 3
  - determinista, 3
- Fórmula
- s de derivación, 312
  - s de integración, 312
  - s para exponentes, 310
  - s para logaritmos, 310
  - s para sumas, 311
  - de inclusión-exclusión, 41
  - de Stirling, 199
- Función
- beta, 215
  - de acumulación conjunta, 253
  - de dist.  $N(0, 1)$ , 224
  - de distribución, 112, 118
  - de probabilidad, 103
  - gama, 211
  - indicadora, 171

- signo, 141
- Función de densidad
  - condicional, 268
  - conjunta, 246
  - de un vector, 246
- Función de distribución
  - bivariada, 254
  - condicional, 270
  - conjunta, 253
  - de un vector, 253
  - marginal, 262
- Función de probabilidad, 103
  - condicional, 268
  - conjunta, 243
  - marginal, 258
- Función
  - de probabilidad
    - simétrica, 151
  - indicadora, 15, 103
- Función de probabilidad
  - condicional, 111
- Función generadora
  - de momentos, 160
  - de probabilidad, 155
- Imagen inversa, 96
  - definición, 102
  - propiedades, 102
- Independencia
  - condicional, 86
  - de eventos, 79
  - de variables aleatorias, 126, 264
  - de varios eventos, 81
- Kolmogorov, A. N., 92
- Laplace, P.-S., 90
- Lema de Abel, 313
- Ley de los grandes números, 288
- Leyes de De Morgan, 10
- Media, 131
- Mediana, 153
- Medida de probabilidad, 33
  - continuidad, 87
  - otras propiedades, 39
- Momento
  - s, 149
  - s absolutos, 150
  - s absolutos centrales, 150
  - s centrales, 150
  - s generalizado, 150
- Notación *o* pequeña, 313
- Ordenaciones
  - con repetición, 51
  - sin repetición, 52
- Paradoja
  - de San Petersburgo, 140
- Permutaciones, 53
- Potencia de un conjunto, 11
- Principio de multiplicación, 50
- Probabilidad, 33
  - axiomática, 32
  - clásica, 18
  - condicional, 63
  - conjunta, 243
  - de Laplace, 19
  - de un evento, 18
  - frecuentista, 28, 291
  - geométrica, 21
  - marginal, 258
  - subjettiva, 31

- Problema de
  - cumpleaños, 58
- Producto
  - Cartesiano, 12
- Sigma álgebra
  - de Borel, 45
- Stirling, 199
- Suceso, 5
- Teorema
  - central del límite, 295
  - de Bayes, 73
  - de cambio de variable, 123
  - de equipartición asintótica, 294
  - de probabilidad total, 68, 69
  - del binomio, 54
    - extensión, 61
  - del estadístico inconsciente, 133
  - multinomial, 56
- Triángulo de Pascal, 54
- Urna de Polya, 67
- Valor
  - esperado, 131
  - promedio, 131
- Variable aleatoria, 95
  - continua, 100
  - discreta, 100
  - distribución de una, 119
- Varianza, 142
  - propiedades, 144
- Vector aleatorio, 241
  - continuo, 241
  - discreto, 241